

FUNDAÇÃO UNIVERSIDADE FEDERAL DE RONDÔNIA
CAMPUS JI-PARANÁ DEPARTAMENTO DE MATEMÁTICA E
ESTATÍSTICA

TUANNHY ROZEIRA HAVERROTH

ANÁLISE FATORIAL PARA IDENTIFICAR FATORES QUE INFLUENCIAM NA
DEMANDA DE ANÁLISES DE DADOS EXTERNOS E INTERNOS NAS EMPRESAS
DE JI-PARANÁ/RO.

Ji-Paraná-RO

2017

TUANNHY ROZEIRA HAVERROTH

**ANÁLISE FATORIAL PARA IDENTIFICAR FATORES QUE INFLUENCIAM NA
DEMANDA DE ANÁLISES DE DADOS EXTERNOS E INTERNOS NAS EMPRESAS
DE JI-PARANÁ/RO.**

Trabalho de Conclusão de Curso apresentado
ao Departamento de Matemática e Estatística,
da Fundação Universidade Federal de
Rondônia (UNIR) *Campus* de Ji- Paraná como
parte dos requisitos para obtenção do título de
Bacharel em Estatística.

Orientador: Dr. Nerio Aparecido Cardoso

Ji-Paraná-RO

2017

Dados Internacionais de Catalogação na Publicação
Fundação Universidade Federal de Rondônia
Gerada automaticamente mediante informações fornecidas pelo(a) autor(a)

H387a Haverroth, Tuanny Rozeira.

Análise Fatorial para identificar fatores que influenciam na demanda de análises de dados externos e internos nas empresas de Ji-Paraná/RO. / Tuanny Rozeira Haverroth. -- Ji-Paraná, RO, 2017.

68 f. : il.

Orientador(a): Prof. Dr. Nerio Aparecido Cardoso

Trabalho de Conclusão de Curso (Graduação em Estatística) - Fundação Universidade Federal de Rondônia

1. Análise de dados. 2. Pesquisa de Mercado. 3. Inteligência de Mercado. 4. Estatística nas empresas. 5. Análise Fatorial Exploratória. I. Cardoso, Nerio Aparecido. II. Título.

CDU 519.237

Bibliotecário(a) Alex Almeida

CRB 11.853

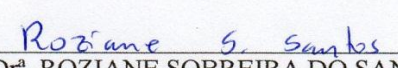


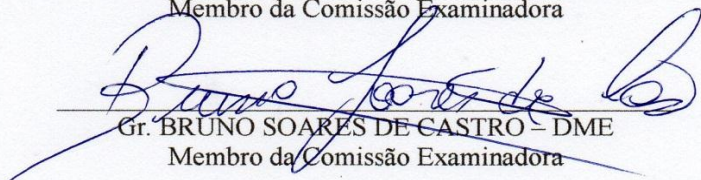
MINISTÉRIO DA EDUCAÇÃO
FUNDAÇÃO UNIVERSIDADE FEDERAL DE RONDÔNIA
Campus de Ji-Paraná - RO
Departamento de Matemática e Estatística

ATA DE DEFESA DO TRABALHO DE CONCLUSÃO CURSO

Ata dos Trabalhos da Comissão Examinadora do Trabalho de Conclusão de Curso da estudante **TUANNHY ROZEIRA HAVERROTH** como parte dos requisitos para obtenção do título de Bacharel em Estatística pela Fundação Universidade Federal de Rondônia – UNIR, orientada pelo Prof. Dr. Nerio Aparecido Cardoso. Integraram a Comissão os Professores Departamento de Matemática e Estatística, Dr. Nerio Aparecido Cardoso, Dr^a. Roziane Sobreira do Santos e Gr. Bruno Soares De Castro. Aos 8 dias do mês de dezembro de 2017 às: 14:40 horas, na sala do Laboratório de Estatística - LABEST do Curso de Bacharelado em Estatística do Departamento de Matemática e Estatística – Campus da UNIR de Ji-Paraná, realizou-se a apresentação pública do trabalho de Conclusão de Curso pelo estudante. O presidente abriu a sessão agradecendo a participação dos membros da Comissão Examinadora. Em seguida convidou a estudante para que fizesse a exposição do trabalho intitulado: “**ANÁLISE FATORIAL PARA IDENTIFICAR FATORES QUE INFLUÊNCIAM NA DEMANDA DE ANÁLISES DE DADOS EXTERNOS E INTERNOS NAS EMPRESAS DE JI-PARANÁ/RO**”. Finalizada a apresentação, cada membro da Comissão Examinadora realizou a arguição para com a estudante. Dando continuidade aos trabalhos, a comissão declara **APROVADA** com conceito **100**. Em seguida, deu por encerrada a reunião, da qual eu, Prof. Nerio Aparecido Cardoso, professor orientador, lavrei a presente ata que vai assinada por mim e pelos membros da Comissão Examinadora.


Dr. NERIO APARECIDO CARDOSO – DME
Membro Presidente da Comissão Examinadora


Dr^a. ROZIANE SOBREIRA DO SANTOS – DME
Membro da Comissão Examinadora


Gr. BRUNO SOARES DE CASTRO – DME
Membro da Comissão Examinadora

Dedico este trabalho, com todo meu amor, primeiramente a Deus Jesus e a Virgem Maria, a todos os seres de luz que me guardam, guiam e me iluminam em meus caminhos. A minha mãe Jucélia, meu pai Dejanir e a todos os meus irmãos, familiares e amigos, pessoas que eu amo respeito e honro.

AGRADECIMENTOS

Primeiramente a Deus, onde busco força e sabedoria para trilhar minha caminhada.

Em especial a minha mãe Jucélia que me deu a vida e educação, não só de estudo, mas principalmente de valores. Por sempre me apoiar, amar e pela dedicação em todos os momentos de minha vida.

Ao meu pai Dejanir que sempre acreditou em mim, como pessoa e profissional. Por todas as oportunidades que me proporcionou.

Aos meus irmãos, amigos e familiares que acreditaram em mim e pela compreensão dos momentos em que não pude estar presente. Ao meu irmão Paulo que nos dias que não pude comparecer ao trabalho para estudar para minhas provas, ele me ajudava no trabalho além de me auxiliar em alguns trabalhos de pesquisa em campo para a universidade.

Aos professores Roziane, Dilson, Jossivan, Lenilson, Enoque, Fanny, Marlos e aos demais professores do Departamento de Matemática e Estatística pelos ensinamentos, amizade, motivação, conhecimento, paciência, compreensão, dedicação e incentivo para eu continuar seguindo no caminho do conhecimento.

Ao professor Bruno Soares pelos ensinamentos e principalmente por me incentivar a ingressar no curso de Estatística, sem seu incentivo não chegaria até aqui.

Em especial ao professor Nerio Cardoso, meu orientador pela dedicação, paciência, incentivo, motivação, amizade e ensinamentos. Por tornar este trabalho possível. Por acreditar em mim e na minha capacidade.

Aos meus amigos de turma por todo companheirismo e amizade durante esses anos, em especial Aline Marina, primeira amizade que cativei dentro da Universidade e os sobreviventes Luís e Pablo, que chegaram até o fim do curso junto comigo, por sempre me auxiliarem quando precisei.

Aos meus irmãos, amigos e Mestres da Barquinha de Ji-Paraná e MDV de Vilhena pelo carinho, pelo acolhimento, pelos momentos de alegrias, união e companheirismo.

Agradeço a todas as pessoas que passaram e estão na minha vida, por me ajudarem a construir quem sou, pelo apoio nos momentos em que mais precisei. Amo vocês!

***"Alguns usam a Estatística como os bêbados usam o poste:
mais para apoio do que para iluminação."***

(Andrew Lang)

ANÁLISE FATORIAL PARA IDENTIFICAR FATORES QUE INFLUENCIAM NA DEMANDA DE ANÁLISES DE DADOS EXTERNOS E INTERNOS NAS EMPRESAS DE JI-PARANÁ/RO.

RESUMO: O presente trabalho de investigação científica permitiu verificar se os empresários de Ji-Paraná/RO utilizam análises de dados internos e/ou externos em seus empreendimentos. E para os que responderam que não utilizam nenhuma análise de dados, identificou-se o porquê de não utilizarem as análises para se orientarem nas tomadas de decisões e amenizar risco, e assim descobrir se o empreendedorismo de Rondônia esta sendo embasado na utilização de análise de dados e se o mercado vem absorvendo os profissionais dessa área, principalmente os estatísticos. Considerando uma amostra de 196 empresários, foi adotada a técnica estatística multivariada de Análise Fatorial Exploratória para declarar os fatores, muito comum em estudos que se utilizam uma escala de avaliação de várias variáveis, pois possibilita reduzir em uma quantidade menor denominadas de fatores as variáveis avaliadas. A Análise Fatorial atribuída da rotação Varimax identificou os fatores que influenciam na adoção de conhecimentos estatísticos na gestão das empresas de Ji-Paraná, sendo eles: Benefícios, Dificuldades, Inovação e Organização.

Palavras-Chave: Análise de dados, Pesquisa de Mercado, Inteligência de Mercado, Estatística nas empresas, Análise Fatorial Exploratória.

LISTA DE TABELAS

Tabela 1 - Classificação da Aplicação do Coeficiente Alpha de Cronbach.	21
Tabela 2 - Classificação da Aplicação da Análise Fatorial pela Estatística KMO.....	26
Tabela 3 - Classificação da pesquisa realizada em 196 empresas de Ji-Paraná em 2017.	33
Tabela 4 - Descrição das 25 variáveis avaliadas pelos empresários sobre a utilização de análise de dados internos e externos nas empresas de Ji-Paraná/RO.	34
Tabela 5 - Quantidade de empresas ativas em Ji-Paraná, segundo seu porte econômico, de acordo com o SEBRAE.	35
Tabela 6 - Comparativo da proporção de Empresas ativas vs. População, no estado de Rondônia em 2015 de acordo com o IBGE.....	36
Tabela 7 - Número de empresas abertas em Ji-Paraná nos últimos três anos	36
Tabela 8 - Empresas entrevistadas em Ji-Paraná/2017, por meio de uma pesquisa exploratória de acordo com os critérios de classificação do porte econômico, utilizado pelo SEBRAE (tamanho amostral = 196).....	38
Tabela 9 - Resultado do teste de normalidade de Shapiro-Wilk da pesquisa com empresários, realizada em Ji-Paraná em 196 empresas.	40
Tabela 10 - Medida de adequação de cada variável (I)	42
Tabela 11 - Medida de adequação de cada variável (II).....	44
Tabela 12 - Matriz Fatorial não rotacionada (I).	46
Tabela 13 - Medida de adequação de cada variável (III).	48
Tabela 14 - Matriz Fatorial não rotacionada II.....	50
Tabela 15 - Rotação ortogonal da matriz fatorial – Varimax.....	51
Tabela 16 - Interpretação dos fatores a partir do critério de rotação Varimax.....	54

LISTA DE FIGURAS

Figura 1 - Setor de atuação das 196 empresas de Ji-Paraná entrevistadas por meio de pesquisa exploratória em 2017.....	37
Figura 2 - Percentual das 196 empresas entrevistadas que investem em análise	38
Figura 3 - Matriz de Correlação das 25 variáveis.....	41
Figura 4 - Matriz de Correlação das 22 variáveis que permanecem no estudo.	43
Figura 5 - Teste Scree-plot para definir número de fatores (I).....	45
Figura 6 - Matriz de Correlação das 12 variáveis que permanecem no estudo.	47
Figura 7 - Teste Scree-plot para definir número de fatores (II).....	49
Figura 8 - Visualização da Variância Total.	52
Figura 9 - Variáveis latentes extraídas do agrupamento das variáveis através da AF.....	53
Figura 10 - Comparativo dos gráficos de dispersão dos Scores nos Planos Fatores 1-2, 1-3 e 1-4 com a variável categórica “Setor”.	56
Figura 11 - Comparativo dos gráficos de dispersão dos Scores nos Planos Fatores 1-2, 1-3 e 1-4 com a variável categórica “Já contratou”.	57

SUMÁRIO

1. INTRODUÇÃO	12
2. JUSTIFICATIVA	13
3. OBJETIVOS.....	13
3.1 Objetivo Geral.....	13
3.2 Objetivos Específicos.....	14
4. FUNDAMENTAÇÃO TEÓRICA	14
4.1 Terminologias para análise de dados empresariais	14
4.1.1 Análises de dados externos	14
4.1.2 Análises de dados internos.....	17
4.2 – Análise Fatorial	18
4.2.1 Problema da pesquisa e definição da amostra.....	20
4.2.2 Verificação de pressupostos da Análise Fatorial	21
4.2.3 Modelo de fatores ortogonal	26
4.2.4 Estimação do número de fatores	29
4.2.5 Estimação das matrizes dos fatores pelo método componentes principais	29
4.2.6 Rotação Ortogonal dos Fatores	30
4.2.7 Seleção das Variáveis Representativas.....	32
4.2.8 Validação do Modelo.....	32
5. METODOLOGIA.....	32
6. RESULTADOS E DISCUSSÕES	35
6.1 Análise descritiva do perfil das empresas de Ji-Paraná.....	35
6.2 Aplicação da Análise Fatorial nos dados.....	39
6.2.1 Verificação da eficiência do questionário.....	39
6.2.2 Normalidade.....	39
6.2.3 Testando a Correlação	40
6.2.4 Definindo o número de fatores	44
6.2.5 Extração dos fatores pelo método de componente principal	49
6.2.6 Rotação Ortogonal da matriz fatorial - Varimax.....	50
6.2.7 Interpretação dos Fatores	52
6.2.8 Validação dos resultados.....	57
7. CONCLUSÕES	59
8. PROPOSIÇÕES.....	60
REFERÊNCIAS BIBLIOGRÁFICAS	61
APÊNDICES	63

1. INTRODUÇÃO

Atualmente uma das ciências mais utilizadas em todas as áreas do conhecimento é a Estatística, que dispõe de uma gama de técnicas estatísticas para analisar dados para tomada de decisão diante de situações incertas, ou seja, extrair informação dos dados para obter uma melhor compreensão das situações em que se está inserida.

As análises estatísticas são utilizadas em diferentes contextos, como pesquisas eleitorais, saúde como em controle de qualidade de doenças, sociais como taxas populacionais, índices de desenvolvimento, índices de desemprego, análises de crescimento, na economia, etc.

Nos negócios as análises estatísticas vêm sendo muito utilizadas nos diversos setores empresariais como no financeiro, marketing, satisfação e perfil de clientes, etc. Porém as análises estatísticas de dados são soluções negligenciada, dentro das micro e pequenas empresas no interior do país, por seus gestores que seguem trabalhando sem conhecer seu público-alvo, fornecedores, custos fixos e variáveis, concorrência, etc.

No relatório Global Entrepreneurship Monitor (GEM) – Monitoramento Global de Empreendedorismo, realizado em 62 países no ciclo de 2015 da pesquisa, o Brasil atingiu o maior nível de empreendedorismo da história do país, alcançando três recordes. Sendo eles, a cada dez brasileiros, quatro possuem um negócio ou criaram um negócio próprio em 2015. A maior Taxa de Empreendedores Iniciais; e a maior Taxa de Empreendedores Estabelecidos.

Entretanto apesar de todo esse movimento de incentivo ao empreendedorismo no país e no Estado, a realidade é que antes de abrir uma empresa, muitos empreendedores não procuram saber o que os clientes e o mercado querem, simplesmente começam a vender o que lhes parece interessante sem ao menos saber se de fato esse produto/serviço oferecido atende aos interesses de alguém ou até mesmo se a ideia irá atrair consumidores.

Segundo o IBGE (Instituto Brasileiro de Geografia e Estatística), em 2010 de cada cem empresas abertas no Brasil, 48 encerraram suas atividades em três anos.

Em divulgação no site oficial do SEBRAE, serviço de apoio à micro e pequena empresa, entre as principais razões para a mortalidade precoce das empresas estão na falta de gestão, de planejamento, informações sobre o mercado em que deseja atuar, desconhecimento do público alvo, quais são as práticas dos concorrentes e tomadas decisões as “cegas”, ou seja, sem parâmetros reais e confiáveis.

A principal ação que o SEBRAE adotou ao longo dos anos, para ajudar na consolidação das empresas no mercado, foi à conscientização da importância de se introduzir na cultura empresarial a realização de pesquisa de mercado nas empresas, bem como a Análise de dados internos para auxiliar e embasar as tomadas de decisões.

Portanto, a grande questão é, se as análises de dados internos ou externos são ferramentas tão úteis para as empresas e principalmente para sobrevivência dos empreendimentos no mercado, por que as empresas e gestores das mesmas não as utilizam?

Em meio a este paradoxo entre o crescimento do empreendedorismo no país, principalmente em Rondônia, e a curta vida útil dos empreendimentos por falta de planejamento e o não uso de pesquisa de mercado e análise de dados para tomada de decisões. O presente estudo científico permitiu analisar se os empresários de Ji-Paraná/RO utilizam análise de dados externos e internos. Com o objetivo de definir indicadores que contribui para que os empresários adotarem análises estatísticas na gestão das empresas e com isso descobrir se o empreendedorismo de Ji-Paraná/RO esta sendo embasado pela utilização de pesquisas de mercado e análise de dados.

A identificação e análise dos indicadores se deu pela técnica estatística multivariada Análise Fatorial Exploratória, justificado por ser uma das técnicas estatísticas de redução de dados a partir de um número volumoso de variáveis independentes.

2. JUSTIFICATIVA

O presente estudo justifica-se por ser um futuro instrumento para criação de estratégias que possam auxiliar a difundir a utilização de análise de dados internos e externos, de acordo com a realidade das empresas, no município e por consequência gerar uma demanda de emprego para os profissionais que se formam anualmente no curso de Bacharelado em Estatística da Fundação Universidade Federal de Rondônia de Ji-Paraná.

3. OBJETIVOS

3.1 OBJETIVO GERAL

Este estudo tem o objetivo de identificar indicadores relevantes que define a demanda de análises de dados internos e externos nas empresas de Ji-Paraná/RO para tomada de decisões.

3.2 OBJETIVOS ESPECÍFICOS

- Identificar o grau de conhecimento dos empresários sobre a importância da utilização de análise de dados dentro das empresas;
- Verificar pontos que impedem as empresas de investirem em análise de dados para extrair informações úteis para suas tomadas de decisões.
- Verificar o capital de investimento das empresas de Ji-Paraná em análise de dados;
- Analisar a relação do faturamento da empresa e a efetiva contratação desse serviço;
- Identificar se as empresas de Ji-Paraná armazenam dados;
- Verificar se as empresas adotam análise de dados para tomada de decisões.

4. FUNDAMENTAÇÃO TEÓRICA

4.1 TERMINOLOGIAS PARA ANÁLISE DE DADOS EMPRESARIAIS

Análise de dados empresariais se resume ao ato de pesquisar, buscar informação que é tão inerente ao homem quanto seus cinco sentidos. Faz parte do desenvolvimento da sociedade como um todo buscar informação, para evoluir (ABEP, 2017).

Segundo o Dicionário Aurélio a definição de Pesquisa é: “*Indagação, investigação* ou busca minuciosa para a averiguação da realidade”.

Já a Associação Americana de Marketing (American Marketing Association) define a pesquisa direcionada ao mercado, as empresas e aos negócios como o objetivo de coletar, registrar e analisar todos os fatos referentes aos problemas relacionados à transferência e venda de mercadorias e serviços do produto ao consumidor.

É possível coletar dados confiáveis para análises de duas maneiras, a coleta de dados externa e a coleta de dados interna. Entretanto com o avanço da tecnologia e aperfeiçoamento das técnicas de análises, dependendo do setor da empresa, que será aplicado essas pesquisas, ou até mesmo por questões regionais, os nomes utilizados para se referir a essas análises de dados interno e externos mudam. Sendo assim é importante esclarecer a diferença de uma da outra e quando utilizá-las.

4.1.1 ANÁLISES DE DADOS EXTERNOS

Análise de dados externos no contexto empresarial se refere ao ato de buscar dados fora da empresa, para analisá-los e transformar em informações relevantes para tomada de decisões (RODRIGUES e BLATTMANN, 2014).

A busca por dados externos, ou seja, fora das empresas é conhecida geralmente por pesquisa de mercado e opinião que evoluiu muito nos últimos 70 anos de história. No Brasil, começou durante a década de 1910 com o crescimento da atividade industrial e o desdobramento do capital vindo do café e com a chegada das empresas americanas que junto vieram também suas agências de propaganda. A primeira década do século XXI foi de consolidação dos grandes grupos globais de pesquisa de mercado no Brasil. Gradativamente as empresas que apresentaram maior crescimento durante a década de 90 foram sendo adquiridas por estes grupos globais que passaram a ditar a forma de funcionamento do mercado (ABEP, 2017).

Ibope, Nielsen, Ipsos, GfK, TNS RI, Millward Brown e Synovate - responde por cerca de 75% do mercado brasileiro de pesquisa de mercado. Outros grupos globais continuam a virem para o Brasil, montando operações ou adquirindo empresas locais, mas a dinâmica de funcionamento do mercado local, em sintonia com o mercado global, já está estabelecida (ABEP, 2017).

De acordo com a Associação Brasileira de Empresas de Pesquisa – ABEP, a primeira pesquisa de mercado que se tem registro no Brasil foi realizada pela N. W. Ayer & Son em 1934. Como objetivo estudar os hábitos dos consumidores de café, contratado pelo Departamento Nacional do Café, para entender as razões da queda no consumo de café para subsidiar o planejamento de campanhas publicitárias.

Atualmente a um mercado de pesquisas e análise de dados é muito forte, em sintonia com os movimentos globais, empresas fortemente consolidadas, organizadas em uma associação reconhecida no mercado e profissionais respeitados no Brasil e no mundo. (ABEP, 2017).

Os termos mais usuais para referir-se a análise de dados externos são (RODRIGUES e BLATTMANN, 2014):

- a) **Estatísticas nacionais:** Regularmente obtidos e organizados por entidades públicas ou privadas. Que apresentam dados econômicos como a cotação de moedas estrangeiras, inflação do país, aumento ou diminuição dos valores de determinadas mercadorias, previsão para as vendas de um dado setor econômico. Estatísticas sociais, econômicas e financeiras disponibilizadas aberta e gratuitamente.
- b) **Técnicas de observação/Boca-Boca:** O pesquisador observa o fenômeno social no seu ambiente natural. As observações podem ocorrer transversalmente (observações

feitas de uma só vez) ou longitudinalmente (observações que ocorrem ao longo de determinados períodos). Exemplos são a análise do uso de produtos e a utilização de cookies para observar comportamento na internet.

- c) **Pesquisa de Mercado:** É a coleta de dados junto ao consumidor, concorrentes ou fornecedores pra orientar a tomada de decisões ou solucionar problemas.
- d) **Pesquisa de Opinião:** Foca nos assuntos sociais e políticos, com tudo também é utilizado na área empresarial.
- e) **Pesquisas Comerciais:** Com foco em identificação de novas praças para atuar. Dimensionar o mercado, identificar o segmento mais lucrativo, detectar novas tendências, identificar a quantidade ou volume que o mercado é capaz de absorver e a que preços esses produtos poderão ser vendidos.
- f) **Pesquisa de segmentação de Mercado:** É o estudo que proporciona a identificação de características de compradores, níveis de demanda, efetividade dos canais de distribuição de produtos e a participação da empresa no mercado, De forma geral, esse estudo é um levantamento de informações estratégicas que podem auxiliar no gerenciamento da empresa.
- g) **Pesquisa de satisfação:** Conhecer o perfil do cliente: ela fornece a caracterização dos clientes nos aspectos quantitativos (potencial do mercado, participação da empresa no mercado, etc.) e qualitativos (estilo de vida, características comportamentais, hábitos de consumo, escolaridade, renda, dentre outros).
- h) **Pesquisas de Clima Organizacional:** Coleta de dados junto aos colaboradores em relação aos diversos fatores que afetam os níveis de motivação e desempenho dos mesmos.
- i) **Pesquisa de Concorrentes:** Observar os concorrentes referentes aos seus pontos fortes e fracos, sistema de vendas e distribuição, políticas de preços e cobrança, qualidade dos produtos e serviços, analisar os fornecedores, bem como o desempenho de seus produtos e serviços.
- j) **Enquetes:** São pesquisas instantâneas de apenas uma pergunta e algumas alternativas, que visam medir a opinião do grupo sobre determinada questão. É típico de pesquisas da internet.

É importante destacar que o objetivo do trabalho científico apresentado não é considerar o nome/conceito específico utilizado ao se referir à análise de dados externos. Mas

evidenciar o conhecimento que os empresários de Ji-Paraná têm sobre o sentido em si a respeito do que é e a utilidade das análises extraídas de uma coleta de dados externa, independente do nome que eles conhecem ou costumam utilizar.

4.1.2 ANÁLISES DE DADOS INTERNOS

Nos últimos 17 anos testemunhou-se um avanço tecnológico estrondoso derivado de investimentos na infraestrutura de negócios que têm melhorado a capacidade de coletar dados a todo o momento em todos os setores das empresas. Praticamente todos os aspectos dos negócios estão em constante coleta de dados. Este amplo armazenamento e disponibilidade de dados levou a necessidade de se extrair informações úteis e conhecimento a partir de grande volume de dados, sendo considerado o novo passo da era digital (PROVOST e FAWCETT, 2016).

Esta necessidade de se extrair dados a partir de um grande volume de dados armazenados (Big Data), foi o que levou a profissão de estatístico como a melhor de 2017 nos Estados Unidos, segundo estudo recente disponível no site da CareerCas (CAREERCAST, 2017), sendo a profissão mais requisitada pelas empresas nos Estados Unidos e com uma promessa de explosão no Brasil. Isso por que o estatístico é o profissional que possui conhecimento de distintas técnicas e métodos focados na construção de modelos e na extração de informações, o que leva as empresas a obterem vantagens competitivas a partir do *Data Science*, mais conhecido como a Ciência dos Dados, que nada mais é do que o “Conjunto de princípios fundamentais que norteiam a extração de conhecimento a partir de dados” como definiu (PROVOST e FAWCETT, 2016).

No livro “O Poder do Hábito: Por que fazemos o que fazemos na vida e nos negócios”, (DUHIGG, 2012) trás uma coleção de relatos reais de empresas norte americanas que investiram na análise de dados internos para embasar suas estratégias e tomada de decisões e a se sobressaírem em relação à concorrência, trazendo inovação e lucro para seus negócios.

Um dos casos de benefício de decisão orientada por dados, retratado no livro, mais conhecido foi o da empresa Target que desejava sair na frente da concorrência. Eles queriam saber se conseguiriam prever quais clientes mulheres estavam esperando bebê antes delas começarem a comprar produtos de maternidade. Isso foi possível de acordo com os históricos de hábitos de consumo de mulheres que souberam posteriormente que estavam grávidas. Mulheres grávidas costumam mudar a dieta, o guarda-roupa, as vitaminas. Esses indicadores foram extraídos de dados históricos, tudo o que o cliente já comprou, dos cartões de créditos,

ligações, movimentação em redes sociais, buscas na internet. Todos esses dados foram montados em modelos preditivos e em seguida implantados em campanhas de marketing que enviavam cupons às futuras mães com descontos em fraldas ou qualquer outro produto ligado à gravidez. Em relação à eficácia do modelo preditivo a Target obteve sucesso, no entanto alguns clientes não gostaram muito da prática, mas os resultados da companhia mostram que as análises geram bons dividendos. A receita da Target cresceu após a aplicação do Data Science cerca de U\$ 44 bilhões em 2002 para US\$ 67 bilhões em 2010.

Com o avanço da tecnologia aliada a Estatística a análise de dados passou a ser chamada por distintos nomes, muitas vezes em inglês, o que pode ser uma dificuldade para as pequenas empresas que não conhecem o que significa e acabam acreditando que por ter nomes tão complicados são coisas muito distantes de sua realidade. Para dirimir essa crença vejamos alguns nomes utilizados hoje para se referir à análises extraída de dados internos, ou seja, dados produzidos diariamente pelas empresas. Como por exemplo, cadastro de cliente, registro de compra e venda de mercadoria, fluxo de caixa, etc.

- a) **Business-intelligence/Inteligência de Negócios:** Forma de analisar um conjunto amplo de dados gerados pela própria empresa, aplicando apenas uma estatística descritiva dos dados (PROOF, 2015).
- b) **Business-analytics/ Análises de Negócios:** Possuem mais recursos de estatísticas, criação de modelos prescritivos e preditivos (PROOF, 2015).
- c) **Data Science/ Ciência dos dados:** Conjunto de princípios fundamentais que norteiam a extração de conhecimento a partir de dados (PROVOST e FAWCETT, 2016).
- d) **Data Mining/Mineração de dados:** É a extração de conhecimento a partir do Data Science, por meio de tecnologia que incorporam esses princípios (SILVA, PERES e BOSCAROLI, 2016).

4.2 – ANÁLISE FATORIAL

O desenvolvimento inicial da análise Fatorial deve-se a Charles Spearman que no início dos anos 1900, procurou descrever a inteligência através de um único fator “g”. Entretanto apenas em 1931 o pesquisador Louis L. Thurstone criou o termo “Análise Fatorial”. Thurstone também identificou “habilidades mentais primárias”, em vez de um único fator “g”, como sugeriu Spearman. A questão é que tal “habilidade” foi um passo importante

para o desenvolvimento da Análise Fatorial, que atualmente é utilizada em várias áreas do conhecimento para diferentes fins.

A Análise Fatorial possui muitas definições, no geral é uma técnica multivariada aplicada a uma série de variáveis denominadas fatores onde se analisa os padrões de correlação entre todas as respostas dadas por todos os pesquisados. Ao realizar análise de correlação entre as respostas de forma a agrupá-las em uma quantidade menor de fatores comparada a quantidade inicial (HAIR, BLACK e BABIN, 2009).

Em Estatística essa série de perguntas denomina-se variáveis, cada uma delas se refere à determinada característica dos elementos da amostra que nos interessa averiguar estatisticamente. Segundo (MANLY, 2008), existem distintos tipos de variáveis, que precisam ser muito bem observadas e definidas para a aplicação da técnica multivariada de Análise Fatorial. Pois a AF só pode ser aplicada em variáveis qualitativa ordinal e variáveis quantitativas contínua e discreta.

Vejamos a seguir a definição dos tipos de variáveis.

- Variável Qualitativa Ordinal: São aquelas que podem ser colocadas em ordem, por exemplo, a classe social (A, B, C, D, ou E), ou sequência de avaliação (ótimo, bom, regular, ruim e péssimo).
- Variável Qualitativa Nominal: São aquelas que não podem ser hierarquizadas ou ordenadas, não tem nenhuma ordem de variações, como a cor dos olhos, o local de nascimento, sexo, carreira, região onde mora. Esta é o único tipo de variável que não é possível aplicar a Análise Fatorial.
- Variável Quantitativa Discreta: O conjunto de resultados possíveis pode ser finito ou enumerável. Exemplo: número de filhos, quantidade de pessoas e etc.
- Variável Quantitativa Contínua: Os valores formam um intervalo ou união de números reais. Exemplo: peso, massa, altura, pressão sistólica, nível de açúcar no sangue.

A Análise Fatorial é frequentemente utilizada em variáveis qualitativas ordinais, como por exemplo, em escalas de avaliações como do tipo de pesquisas sobre a satisfação do cliente, quando os clientes avaliam uma grande quantidade de atributos do negócio.

Os objetivos da Análise Fatorial podem ser divididos de duas formas distintas, AF Exploratória e AF Confirmatória (HAIR, BLACK e BABIN, 2009).

a) Análise Fatorial Exploratória:

Geralmente é utilizada nos estágios mais primários da pesquisa, no sentido de literalmente explorar os dados. Nessa fase, procura-se explorar a relação entre um conjunto de variáveis, identificando padrões de correlação. Podendo também ser utilizada para criar variáveis independentes ou dependentes que podem ser utilizadas posteriormente em modelos de regressão. Para os propósitos deste estudo, foi utilizado Análise Fatorial Exploratória(AFE) (HAIR, BLACK e BABIN, 2009).

b) Análise Fatorial Confirmatória:

Mais utilizado para testar hipóteses. Nesse caso, o pesquisador guiado por alguma teoria testou se determinadas variáveis são representativas de um conceito.

O que diferencia AFE da AFC é que na segunda o usuário indica que estrutura ele imagina existir nos dados e, através da aplicação da técnica, terá fundamentação estatística para afirmar se aquela estrutura é ou não aceitável para explicar o comportamento dos dados.

Independente se a AF for exploratória ou confirmatória a metodologia base é a mesma, onde possuem as seguintes quatro etapas: Definir objetivo da pesquisa; Verificar da adequação da aplicação da análise, calculando a matriz de correlação das variáveis buscando o grau de associação entre as variáveis; Extração dos fatores analisando o poder do modelo na representação dos dados, os mais significativos representarão os dados; Rotação dos fatores para facilitar o entendimento e por fim a geração dos escores fatoriais (HAIR, BLACK e BABIN, 2009).

4.2.1 PROBLEMA DA PESQUISA E DEFINIÇÃO DA AMOSTRA

Primeiramente o pesquisador deve ter atenção ao definir as questões que irão compor seu instrumento de registro, para que o mesmo possa ser estruturado de maneira que ao final responda os objetivos iniciais, lembrando-se que os tipos de variáveis (discretas, ordinais ou contínuas) são determinantes para o sucesso do trabalho.

Os indivíduos que irão compor a população em que as variáveis serão observadas, bem como a quantidade e o critério de amostragem são passos fundamentais. Não foram utilizado método de amostragem probabilística e sim a recomendação citada por (HAIR, BLACK e BABIN, 2009) e (MINGOTI, 2013) de que a amostra de preferência deve ser do maior tamanho possível e bem representativa da população em estudo. Sendo que o tamanho da mostra necessita ter pelo menos 50 observações, tendo como base o cálculo mínimo de cinco vezes mais observações do que o número de variáveis. Sendo o ideal é dez observações para cada variável. Caso as variáveis estejam em unidades de medidas diferentes, convém

padronizar os dados antes de realizar a análise, ou seja, cada variável X_i terá média zero e variância 1.

4.2.2 VERIFICAÇÃO DE PRESSUPOSTOS DA ANÁLISE FATORIAL

A Análise Fatorial como toda técnica estatística possui alguns pressupostos que devem ser atendidos para assegurar a confiabilidade dos resultados. Sendo esses: Verificação da confiabilidade do instrumento de registro; Normalidade, Correlação e Linearidade dos dados; Medidas de adequação da amostra e das variáveis (HAIR, BLACK e BABIN, 2009).

4.2.2.1 INSTRUMENTO DE REGISTRO

Primeiro passo é verificar se o instrumento de registro utilizado para coleta de dados é eficiente para aplicação da técnica AFE. Um teste muito utilizado para tal fim é o Coeficiente Alpha de Cronbach, criado em 1951 por Lee J. Cronbach. Com objetivo de estimar a confiabilidade de um questionário aplicado em uma pesquisa, calculado a partir da variância dos itens individuais e da variância da soma dos itens de um instrumento de registro que utilizem a mesma escala de medição. Podendo medir blocos de questões (construtos, fatores,...) ou o questionário como um todo (MAROCO e MARQUES, 2006).

A fórmula proposta por Cronbach é:

$$\alpha = \frac{k}{k-1} * \left[1 - \frac{\sum_{j=1}^k S_j^2}{S_T^2} \right] \quad (1)$$

Onde k é o número de itens do instrumento, S_j^2 é a variância do item j ($j=1, \dots, k$) e S_T^2 é a variância dos totais da escala definida.

Na tabela 1 apresenta-se a classificação dos resultados da Aplicação do Coeficiente Alpha de Cronbach.

Tabela 1 - Classificação da Aplicação do Coeficiente Alpha de Cronbach.

Valor de α	Confiabilidade
$\alpha > 0,90$	Muito Alta
$0,75 < \alpha \leq 0,90$	Alta
$0,60 < \alpha \leq 0,75$	Moderada
$0,30 < \alpha \leq 0,60$	Baixa
$\alpha \leq 0,30$	Muito Baixa

Fonte: Maroco e Marques, 2006.

4.2.2.2 NORMALIDADE

Os testes de normalidade são utilizados para verificar se a distribuição de probabilidade associada a um determinado conjunto de dados segue a distribuição de Probabilidade de Gauss. Verificar a normalidade dos dados é importante para adiante definir qual melhor método de estimação dos fatores para os dados analisados, veja na seção 4.2.5.

É possível verificar se os dados seguem uma distribuição de normalidade de gráficos e testes estatísticos. É importante verificar a normalidade dos resíduos de ambas as formas, pois o problema não deve ser verificado somente graficamente, uma vez que a decisão de normalidade depende muito da interpretação do pesquisador o que pode variar muito de um para o outro. Graficamente, podem ser feitos gráficos como histogramas e/ou o gráfico normal de probabilidade.

Já os testes têm-se como referência o p-valor. Existem vários testes que podem ser realizados, como por exemplo, os testes de Shapiro-Francia, Cramer-Von-Mises, Anderson-Darling, Kolmogorov-Smirnov, porém por possuir um alto poder de eficiência e ser muito utilizado, adota-se no presente estudo somente o teste de Shapiro-Wilk.

Tendo como hipótese:

H_0 : A amostra provém de uma população normal

H_1 : A amostra não provém de uma população normal

A Estatística do teste de Shapiro-Wilk é dada por W:

$$W = \frac{b^2}{\sum_{i=1}^n (x_{(i)} - \bar{x})^2} \quad (2)$$

Onde que x_i é a variável aleatória observada, \bar{x} a média das variáveis aleatórias observadas e b_i são coeficientes tabelados. A constante b é determinada da seguinte forma:

$$b = \begin{cases} \sum_{i=1}^{n/2} a_{n-i+1} * (x_{(n-i+1)} - x_{(i)}) & \text{se } n \text{ é par} \\ \sum_{i=1}^{(n+1)/2} a_{n-i+1} * (x_{(n-i+1)} - x_{(i)}) & \text{se } n \text{ é ímpar} \end{cases} \quad (3)$$

Em que a_{n-i+1} são constantes geradas pelas médias, variâncias e covariâncias das estatísticas de ordem de uma amostra de tamanho n de uma distribuição Normal encontrados na tabela de distribuição W .

Ao avaliar a estatística do teste através do p -valor define-se primeiramente o nível de significância do teste (α), normalmente 0,05 e como regra de decisão tem-se que se o p -valor $> \alpha=0,05$ não rejeita-se H_0 o que indica que a amostra provém de uma população normal.

4.2.2.3 MODELO FATORIAL VIA MATRIZ DE CORRELAÇÃO

Um dos requisitos da Análise Fatorial é que as variáveis sejam altamente correlacionadas, se estas correlações forem muito baixas não seja apropriado aplicar a Análise Fatorial. Em seguida, serão apresentados os testes de Esfericidade de Bartlett, Medida de adequação da amostra KMO e Medida de adequação de cada variável MSA, que indicam o quanto as variáveis podem estar correlacionadas e assim definir se a matriz de dados está apropriada para adoção da Análise Fatorial.

Seja $X_{p \times 1}$ um vetor aleatório com vetor de médias μ , onde $\mu=(\mu_1 \mu_2 \dots \mu_p)'$, matriz de covariâncias $\Sigma_{p \times p}$ e matriz de correlação $P_{p \times p}$. Sejam $Z_i=[(X_i-\mu_i)/\sigma_i]$ as variáveis originais padronizadas, onde μ_i e σ_i representa respectivamente a média e o desvio padrão da variável X_i , $i=1,2,\dots,p$. Neste caso, a matriz $P_{p \times p}$ é a matriz de covariâncias do vetor aleatório $Z=(Z_1 Z_2 \dots Z_p)'$.

O modelo de Análise Fatorial construído a partir da matriz de correlação teórica $P_{p \times p}$ é um modelo que relaciona linearmente as variáveis padronizadas e os m fatores comuns que são as variáveis latentes que se busca encontrar que a princípio, é desconhecido (MINGOTI, 2013). As equações do modelo são dadas por:

$$\begin{pmatrix} Z_1 = l_{11} F_1 + l_{12} F_2 + \dots l_{1m} F_m + \varepsilon_1 \\ Z_2 = l_{21} F_1 + l_{22} F_2 \dots l_{2m} F_m + \varepsilon_2 \\ \vdots \\ \vdots \\ Z_p = l_{p1} F_1 + l_{p2} F_2 \dots l_{pm} F_m + \varepsilon_p \end{pmatrix} \quad (4)$$

Em notação matricial, o modelo acima pode ser dado por:

$$D(X - \mu) = L F + \varepsilon \quad (5)$$

Onde,

$$(X - \mu)_{PX1} = \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{bmatrix} \quad \varepsilon_{PX1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix} \quad F_{mX1} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_p \end{bmatrix} \quad L_{PXm} = \begin{pmatrix} l_{11} & l_{12} & \dots & l_{1m} \\ l_{21} & l_{22} & \dots & l_{2m} \\ \vdots & \vdots & \dots & \vdots \\ l_{p1} & l_{p2} & \dots & l_{pm} \end{pmatrix}$$

$$D_{PXP} = \begin{bmatrix} 1/\sigma_1 & 0 & 0 & \dots & 0 \\ 0 & 1/\sigma_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1/\sigma_p \end{bmatrix} ; D(X - \mu) = \begin{bmatrix} Z_1 & 0 & 0 & \dots & 0 \\ 0 & Z_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & Z_p \end{bmatrix}$$

O modelo de Análise Fatorial assume que as variáveis Z_i estão relacionadas linearmente com novas variáveis aleatórias $F_j, j = 1, 2, \dots, m$, que será identificadas. Já no modelo F_{mx1} é um vetor aleatório contendo m fatores, que são chamados de variáveis latentes, ou seja, não são observáveis não podendo ser medidos a priori e tem a função de descrever os elementos da população em estudo. O vetor ε_{px1} é um vetor de erros aleatórios e corresponde aos erros de medida e à variação de Z_i que não é explicada pelos fatores comuns $F_j, j = 1, 2, \dots, m$, incluídos no modelo. O coeficiente l_{ij} , comumente chamado de cargas fatoriais, é o coeficiente da i -ésima variável padronizada Z_i e j -ésima o fator F_j e representa o grau de relacionamento linear entre Z_i e $F_j, j = 1, 2, \dots, m$. Os coeficientes l_{ij} ficam armazenados na matriz L_{pxm} que é uma matriz de parâmetros constantes, os quais precisarão ser estimados por algum mecanismo. Assim, de acordo com o modelo linear (4), a informação das p -variáveis originais padronizadas (Z_1, Z_2, \dots, Z_p) estará sendo representada por $(p + m)$ variáveis aleatórias não observáveis, isto é, $(\varepsilon_1 + \varepsilon_2 \dots \varepsilon_p)$ e $(F_1 + F_2 \dots F_p)$. O objetivo é identificar as novas m -variáveis, interpretá-las e calcular seus escores (MINGOTI, 2013).

4.2.2.4 TESTE DE ESFERICIDADE DE BARTLETT

Uma maneira de analisar a matriz de correlação é mediante ao Teste de Esfericidade de Bartlett, que consiste em testar a hipótese de que as variáveis não sejam correlacionadas.

H_0 : É uma matriz de correlação.

H_1 : Não é uma matriz de correlação.

Caso não rejeita-se $H_0 : P = I$, significa que $\rho_{ij} = 0 (i \neq j)$ e a matriz é identidade, o que indica que o modelo fatorial é inapropriado, pois não há correlação entre os dados. Dessa

forma procura-se, com um nível de significância determinado de 5%, rejeitar-se a hipótese nula de matriz de correlação identidade.

A estatística do teste é dada por:

$$X^2 = - \left[(n - 1) - \frac{2p + 5}{6} \right] \ln|R| \quad (6)$$

Que tem uma distribuição qui-quadrada χ^2 de $v = \frac{p(p-1)}{2}$ graus de liberdade.

Onde: n = tamanho da amostra,

p = número de variáveis,

$|R|$ = determinantes da matriz de correlação.

4.2.2.5 MEDIDA DE ADEQUAÇÃO DA AMOSTRA KMO

Para verificar a adequação da amostra é possível adotar o índice de *Kaiser-Meyer-Olkin* (KMO) proposto por Kaiser, Meyer e Olkin. O objetivo do KMO é comparar a amplitude das correlações observadas entre as variáveis com a amplitude das correlações parciais, sendo que esta última indica o grau de relação entre duas variáveis ignorando a influência das outras. O índice KMO esta apresentado a seguir:

$$KMO = \frac{\sum_{j \neq i}^p \sum_{i \neq j}^p \rho_{ij}^2}{\sum_{j \neq i}^p \sum_{i \neq j}^p \rho_{ij}^2 + \sum_{j \neq i}^p \sum_{i \neq j}^p \rho_{ij(p)}^2} \quad (7)$$

Onde em que ρ_{ij} é o coeficiente de correlação entre as variáveis e $\rho_{ij(p)}$ é o coeficiente de correlação parcial.

Observe na tabela 2 a classificação do índice KMO que varia entre 0 e 1, quanto menor as correlações parciais mais o índice se aproxima de 1 (BELFIORE, FÁVERO e ANGELO, 2006).

Tabela 2 - Classificação da Aplicação da Análise Fatorial pela Estatística KMO

KMO	Classificação
0,90 – 1,00	Excelente
0,80 - 0,90	Ótimo
0,70 – 0,80	Bom
0,60 – 0,70	Regular
0,50 – 0,60	Ruim
0,00 – 0,50	Inadequado

Fonte: Belfiore, Fávero e Angelo, 2006.

4.2.2.6 MEDIDA DE ADEQUAÇÃO DE CADA VARIÁVEL MSA

Também é possível dimensionar a medida de adequação para cada variável resposta através da *Measure of Sampling Adequacy* (MSA), de forma similar ao KMO.

A seguir a equação de MSA:

$$MSA = \frac{\sum_{i \neq j}^p \rho_{ij}^2}{\sum_{i \neq j}^p \rho_{ij}^2 + \sum_{i \neq j}^p \rho_{ij(p)}^2} \quad (8)$$

Onde que, ρ_{ij} é o coeficiente de correlação entre as variáveis e $\rho_{ij(p)}$ é o coeficiente de correlação parcial.

Se a estatística MSA apresenta um valor baixo, podemos detectar as variáveis individualmente responsáveis por esse efeito, ou seja, que se encontram no domínio inaceitável e excluí-las (HAIR, BLACK e BABIN, 2009). Quanto maiores forem os valores do MSA, melhor será a utilização da Análise Fatorial. A classificação da análise do MSA é análoga ao KMO, como apresentado na tabela 2.

4.2.3 MODELO DE FATORES ORTOGONAL

Segundo (MINGOTI, 2013) algumas suposições são necessárias para que se possa operacionalizar a estimação do modelo.

(i) Todos os fatores têm média igual à zero, o que implica que.

$$E[F_{mx1}] = E[F_j] = 0, \quad j = 1, 2 \dots, m;$$

(ii) Todos os fatores F_j não são correlacionados e têm variância iguais a 1, ou seja,

$$\text{Var}[F_{mx1}] = I_{mxm} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{bmatrix};$$

(iii) Todos os erros têm médias iguais a zero; o que implica que

$$E[\varepsilon_{px1}] = E[\varepsilon_j] = 0, j = 1, 2, \dots, m;$$

(iv) Erros não são correlacionados entre si e não necessariamente têm a mesma variância, $\text{Var}[\varepsilon_j] = \psi_j$ e $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \forall i \neq j$, ou seja,

$$\text{Var}[\varepsilon_{pxp}] = \psi_{pxp} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & 0 & \psi_p \end{bmatrix};$$

(v) Os vetores do erro ε_{px1} e F_{mx1} são independentes, Portanto a covariância desses vetores é igual à esperança deles que é igual a zero, $\text{Cov}(\varepsilon_{px1}, F_{mx1}) = E(\varepsilon F') = 0$.

A suposição (v) implica dizer que os vetores dos erros e da variância representam duas fontes de variação distintas, relacionadas às variáveis padronizadas. Um modelo fatorial com as suposições (i) e (ii) é chamado de ortogonal, onde a ortogonalidade refere-se ao fato de que os m fatores são ortogonais entre si e por consequência essa suposição está relacionada com a matriz de correlação teórica P_{pxp} (MINGOTI, 2013).

Quando o modelo ortogonal é assumido, a matriz P_{pxp} pode ser reparametrizada na forma:

$$P_{pxp} = L L' + \psi \quad (9)$$

Isso por que:

$$P_{pxp} = \text{Var}(Z) = \text{Var}(LF + \varepsilon) = \text{Var}(LF) + \text{Var}(\varepsilon) = LIL' + \psi$$

Onde I é a matriz identidade de dimensão $_{pxp}$. O objetivo da análise fatorial é encontrar as matrizes L_{pxp} e ψ_{pxp} que possam representar a matriz P_{pxp} para um dado valor de m, menor que o número de variáveis originais p. Existe muitas matrizes de correlação que não podem ser decomposta na forma $LIL' + \psi$ para um valor de m muito menor que p.

$$P_{p \times p} = \begin{bmatrix} \sum_{j=1}^m l_{1j}^2 & \sum_{j=1}^m l_{1j} l_{j2} & \cdots & \sum_{j=1}^m l_{1j} l_{jp} \\ \sum_{j=1}^m l_{2j} l_{j1} & \sum_{j=1}^m l_{2j}^2 & \cdots & \sum_{j=1}^m l_{2j} l_{jp} \\ \vdots & \vdots & \cdots & \vdots \\ \sum_{j=1}^m l_{pj} l_{j1} & \cdots & \cdots & \sum_{j=1}^m l_{pj}^2 \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & 0 & \psi_p \end{bmatrix} \quad (10)$$

Em (10) é possível visualizar melhor o formato das matrizes envolvidas na decomposição dada em (9) da matriz de correlação, onde as decomposições são apresentadas a seguir:

(i₁) $Var(Z_i) = l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2 + \psi_i$, onde $h_i^2 = l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2$, $i = 1, 2, \dots, p$,
O que significa que a variância de Z_i é decomposta em duas partes. A primeira, denotada por h_i^2 , é a variabilidade de Z_i explicada pelos m fatores incluídos no modelo fatorial. Esta parte da variabilidade é chamada de “comunalidade”, nome dado em função do fato de que os fatores F_j , para $j=1, 2, \dots, m$ aparecem em todas as equações do modelo (7) e as variáveis Z_i tem uma fonte de variação em comum. A segunda, denotada por ψ_i , é a parte da variabilidade de Z_i associada apenas ao erro aleatório ε_i , o qual é específico de cada variável Z_i . Esta parte da variabilidade é chamada de “unicidade” ou “variância específica”. Como as variáveis Z_i têm variâncias iguais a 1, segue que $h_i^2 + \psi_i = 1$.

Segundo (HAIR, BLACK e BABIN, 2009) a comunalidade pode ser entendida como a quantidade total de variância que uma variável original compartilha com todas as outras variáveis incluídas na análise. Por isso no estudo de Análise Fatorial os fatores comuns possui maior interesse que os fatores específicos.

Os fatores específicos se adicionam ao modelo por causa da impossibilidade de expressar, em geral, as p -variáveis em função de um número mais reduzido de fatores.

$$(i_2) Cov(Z_i, Z_k) = l_{i1} l_{k1} + l_{i2} l_{k2} \dots + l_{im} l_{km}, \quad i, k = 1, 2, \dots, p, \text{ para } i \neq j.$$

$$(i_3) Cov(Z_i, F_k) = L_{p \times p} \text{ e, portanto, } Cov(Z_i, F_j) = Corr(Z_i, F_j) = l_{ij}, \quad i = 1, 2, \dots, p; \text{ para } j = 1, 2, \dots, m.$$

Isso vem do fato de que:

$$Cov(Z, F) = Cov(LF + \varepsilon, F) = Cov(LF + F) + Cov(\varepsilon, F) = Cov(LF, F) = L$$

Assim, pode-se utilizar a matriz $L_{p \times p}$ na busca do entendimento e interpretação dos fatores F_j , para $j=1, 2, \dots, m$.

Para operacionalizar a análise fatorial na prática, primeiramente, é necessário utilizar mecanismos para estimar o valor de m . A partir do valor estimado de m podemos, então, estimar as matrizes $L_{p \times p}$ e $\psi_{p \times p}$.

(i4) Em relação à variância total, a proporção explicada pelo fator F_j é dada por:

$$PVTE_{FJ} = \frac{\sum_{i=1}^p l_{ij}^2}{P} \quad (11)$$

Sendo que os fatores mais representativos do modelo são aqueles com maiores valores de $PVTE$ (11), segundo autora (MINGOTI, 2013) também é comum expressar seus valores em porcentagem.

4.2.4 ESTIMAÇÃO DO NÚMERO DE FATORES

Os fatores são extraídos na ordem do mais explicativo para o menos explicativo, para a estimação desses fatores basta extrair os autovalores da matriz de correlação amostral. Teoricamente, o número de fatores é sempre igual ao número de variáveis. Entretanto, alguns poucos fatores são responsáveis por grande parte da explicação total. Por tanto observa-se quais autovalores são mais importantes em termos de grandeza numérica, utilizando os seguintes critérios (MINGOTI, 2013).

Critério 1: A análise da proporção da variância total relacionada com cada autovalor. Permanecendo aqueles que representam maiores proporções de variância total.

Critério 2: Critério proposto por Kaiser, define que o valor de m será igual ao número de autovalores maiores ou iguais a 1.

Critério 3: Observação do gráfico scree-plot, que é equivalente ao critério 1. Procura-se no gráfico um ponto de salto, que estaria representando um decréscimo de importância em relação à variância total.

Os critérios descritos levam em consideração apenas a grandeza numérica dos autovalores. No entanto deve-se levar em consideração a interpretabilidade dos fatores e o princípio da parcimônia.

4.2.5 ESTIMAÇÃO DAS MATRIZES DOS FATORES PELO MÉTODO COMPONENTES PRINCIPAIS

Uma vez que a matriz de correlação é adequada e o pesquisador já decidiu o número de fatores a extrair, com base nos métodos apresentados na seção 4.2.4, agora cabe definir o

melhor método de extração dos fatores que esteja consoante ao objetivo da pesquisa e dos dados. A estimação das cargas fatoriais Λ e das variâncias específicas Ψ baseia-se na matriz de covariâncias ou na matriz de correlação se estas variáveis estiverem padronizadas.

Os três métodos mais comuns para estimação são: Método dos Fatores Comuns; Método de Máxima Verossimilhança que exige normalidade e o utilizado no presente trabalho o Método dos Componentes Principais (CP) que não exige normalidade.

O método dos CP possibilita investigações com um grande número de dados disponíveis. Possibilita, também, a identificação das medidas responsáveis pelas maiores variações entre os resultados, sem perdas significativas de informações. Além disso, transforma um conjunto original de variáveis em outro conjunto: os componentes principais (CP) de dimensões equivalentes (HAIR, BLACK e BABIN, 2009).

Essa transformação, em outro conjunto de variáveis, ocorre com a menor perda de informação possível, sendo que está também busca eliminar algumas variáveis originais que possuam pouca informação.

A meta da análise de componentes principais é abordar aspectos como a geração, a seleção e a interpretação das componentes investigadas. Ainda pretende-se determinar as variáveis de maior influência na formação de cada componente, que serão utilizadas para estudos futuros, tais como de controle de qualidade, estudos ambientais, estudos populacionais entre outros.

De acordo com (HAIR, BLACK e BABIN, 2009) o modelo fatorial utilizando o método dos componentes principais é apropriado quando o propósito é a previsão ou o número mínimo de fatores necessários para explicar a parte máxima da variância representada no conjunto original de variáveis, e quando conhecimento anterior sugere que as variâncias específicas representam uma proporção relativamente pequena da variância total.

Uma grande vantagem desse método é que ele não exige a suposição de normalidade das variáveis e sempre proporciona uma solução. Contudo, tem o inconveniente quando as comunalidades forem baixas, particularmente, podendo levar a estimadores viciados contidos na matriz de cargas fatoriais. Maiores informações sobre a estimação de matrizes pode ser encontrado em (MINGOTI, 2013).

4.2.6 ROTAÇÃO ORTOGONAL DOS FATORES

As soluções de fatores não-rotacionados extraem fatores na ordem de sua importância, o primeiro fator tende a ser um fator geral com quase toda a variável com carga significativa, e

explica a quantia maior de variância; o segundo fator e os seguintes são então baseados na quantia residual de variância. Então, o efeito final de rotacionar a matriz fatorial é redistribuir a variância dos primeiros fatores para os últimos com o objetivo de atingir um padrão fatorial mais simples e teoricamente mais significativo (HAIR, BLACK e BABIN, 2009).

Existem vários critérios de rotação, podem ser ortogonais ou oblíquos. Os métodos ortogonais produzem fatores que não estão correlacionados entre si, chamados de fatores ortogonais, sendo interpretados a partir de suas cargas fatoriais. Na rotação oblíqua, os fatores estão correlacionados e, para a interpretação da solução, torna-se necessária a consideração simultânea das correlações e das cargas. É um procedimento de extrema importância para a interpretação dos fatores e assim para a construção das variáveis latentes.

Segundo os autores (MINGOTI, 2013) e (HAIR, BLACK e BABIN, 2009) os métodos rotacionais ortogonais, mais usuais são o Varimax, o Quartimax e o Equamax, descritos a seguir:

- i) **Quartimax:** Rotaciona o fator inicial de modo que uma variável tenha carga alta em um fator e cargas tão baixas quanto possível em todos os outros fatores. Neste método, muitas variáveis podem ter carga alta no mesmo fator, pois o método busca minimizar o número de fatores necessários para explicar uma variável.
- ii) **Varimax:** Minimiza o número de variáveis que têm altas cargas em um fator, assim para cada fator existe apenas alguns pesos significativos e todos os outros sejam próximos de zero. Ou seja, concentra na simplificação das colunas, de modo que cada coluna contenha cargas altas, isto é, próximas a 1 ou -1 e o restantes das cargas dessa coluna sejam próximas a zero.
- iii) **Equamax:** Congrega características dos métodos Quartimax e Varimax, ou seja, seu objetivo é simplificar linhas e colunas simultaneamente (Simplificação dos fatores e das variáveis).
- iv) **Os métodos de rotação oblíqua** mais conhecidos são o Direct Oblimin e Promax, nesses métodos as communalidades são preservadas, porém os fatores gerados apresentam-se mais fortemente correlacionados. Vale destacar que a rotação não afeta a qualidade de ajuste do modelo fatorial, as communalidades e o total da variância explicada pelos fatores. Entretanto, o percentual de variância explicada em cada fator muda após rotação.

4.2.7 SELEÇÃO DAS VARIÁVEIS REPRESENTATIVAS

Se o objetivo do pesquisador é reduzir o número de variáveis para aplicar em estudos posteriores, por meio da análise fatorial é possível selecionar a variável que apresenta maior correlação no fator, que seja mais fácil de mensurar e mais representatividade teórica para representar esse fator. Isso seria uma tarefa simples se em cada fator uma variável apresentar carga bem mais alta que as outras variáveis pertencentes a ele, mas o que ocorre em geral é que duas ou mais variáveis apresenta correlações significantes e próximas. Caso isso ocorra, a seleção da variável representativa dependerá do conhecimento a priori que o pesquisador tem da teoria.

4.2.8 VALIDAÇÃO DO MODELO

De acordo com (HAIR, BLACK e BABIN, 2009) esta etapa de validação do modelo ortogonal fatorial é essencial, particularmente quando se tenta definir uma estrutura latente entre as variáveis, pois reafirma o resultado encontrado a partir da Análise Fatorial. A validação pode ser feita de distintas maneiras, como análise de amostra repartida, aplicação a amostras inteiramente novas ou por generalidade de suas conclusões, sendo esta última a mais utilizada.

A generalidade dos resultados consiste em realizar uma nova análise fatorial a partir dos resultados da análise fatorial anterior aplicando sobre a amostra da mesma população em estudo, se possível. Se não for possível, o pesquisador poderá particionar a amostra original em dois subgrupos aleatórios ou mais e estimar modelos fatoriais para cada um, para isso é necessário que a amostra original tenha um tamanho relativamente grande, em função do número de casos por variável. Por último comparar essas matrizes fatoriais.

5. METODOLOGIA

A identificação e análises dos indicadores ocorreram através da adoção da técnica de Análise Fatorial Exploratória. A coleta dos dados foi realizada no período compreendido entre 30 de julho a 31 de agosto de 2017. Antes do preenchimento dos instrumentos de registros, foi realizado um pré-teste com uma amostra de 15 empresas, a fim de verificar os ajustes necessários ao instrumento de registro, apêndice II. Observou-se que o instrumento era apropriado e não necessitou de ajustes. O instrumento de registro foi elaborado de acordo

com a experiência profissional da pesquisadora, referente análise de dados, junto as empresas do estado de Rondônia.

Observa-se na tabela 3 os aspectos gerais da classificação da pesquisa nas 196 empresas de Ji-Paraná em 2017, bem como seu plano amostral.

Tabela 3 - Classificação da pesquisa realizada em 196 empresas de Ji-Paraná em 2017.

Quanto à/ao:	Método:
Abordagem	Presencial
Procedimentos	Estatísticos
Natureza	Induzido
Forma de abordagem	Qualitativa Ordinal
Análises dos dados	Análise Descritiva e Fatorial
Procedimentos técnicos	Instrumento de registro
Quantidade da Amostra	125 = < X =< 250.

Fonte: Haverroth, 2017.

Antes do trabalho de campo o plano amostral definido foi de no mínimo 125 entrevistas e no máximo 250 entrevistas, segundo a recomendação de (MINGOTI, 2013) e (HAIR, BLACK e BABIN, 2009) apresentada na seção 4.2.1. Pelas circunstâncias e dificuldades dos responsáveis pelas empresas terem disponibilidade de tempo para responder o instrumento de registro. Foram obtidos 196 entrevistas, quantidade suficiente para adotar a técnica de Análise Fatorial.

A coleta de dados ocorreu direta e presencialmente, realizada junto aos empresários ou gerentes das empresas de Ji-Paraná. Foi utilizado como instrumento de pesquisa um documento composto de 30 questões semi-estruturada, dentre elas 5 questões foram para identificação do perfil das empresas utilizando análise descritiva, onde três eram fechadas com espaço para justificativas e duas questões abertas. E 25 afirmações, questão fechada, para realização da Análise Fatorial, onde foi solicitado aos empresários que avaliassem em escala ordinal de Likert “1” a “5” sendo o um “nada relevante” e o cinco “totalmente relevante”, as afirmações descritas na tabela 4 sobre Análise de Dados internos e externos. As variáveis em estudo foram:

Tabela 4 - Descrição das 25 variáveis avaliadas pelos empresários sobre a utilização de análise de dados internos e externos nas empresas de Ji-Paraná/RO.

Variáveis	Afirmações sobre análise de dados internos e externos
Var 1	A empresa faz registro cadastral dos clientes
Var 2	A empresa registra a quantidade e itens de produtos que compra
Var 3	A empresa registra a quantidade e itens de produtos que vende
Var 4	A empresa utiliza os dados dos registros internos para extrair informações para melhorar o planejamento
Var 5	A empresa disponibiliza algum canal de sugestão/reclamação para os clientes
Var 6	Análises de dados ajuda a Conhecer perfil de clientes
Var 7	Entender a necessidade do cliente
Var 8	Auxiliar na tomada de decisões
Var 9	Ajuda na hora de comprar mais mercadoria
Var 10	Criar promoções
Var 11	Conhecer a concorrência
Var 12	Conhecer falhas internas da empresa
Var 13	Aumenta os lucros
Var 14	Melhorar gestão
Var 15	Melhor Direcionar investimentos para propaganda
Var 16	Não fornecem informações importantes
Var 17	Nos resultados não terá nada que eu já não saiba sobre minha empresa e meus clientes
Var 18	Não consigo colocar em prática as informações extraídas das análises.
Var 19	Os valores são altos para contratar esse tipo de serviço
Var 20	Não conheço quem faça esse trabalho
Var 21	Eu mesmo faço minhas análises de forma empírica
Var 22	Temos funcionários capacitados para fazer as análises
Var 23	Difícil contratar um profissional ou empresa dessa área
Var 24	Não gosto de ser avaliado
Var 25	Não gosto de tecnologia

Fonte: Haverroth, 2017.

Devido o fato do instrumento de registro ter sido preenchido direto por pessoas que no caso, foram empresários, donos ou gerentes das empresas de Ji-Paraná/RO. Foi necessário o registro da pesquisa na Plataforma Brasil sistema eletrônico criado pelo Governo Federal para sistematizar o recebimento dos projetos de pesquisa que envolvam seres humanos. O projeto

do presente estudo foi submetido à plataforma no dia 31 de julho e foi aprovado no dia 28 de agosto de 2017.

Após a coleta os dados foram sistematizados em um banco de dados via Software R 3.4.0 onde foi utilizada a técnica estatística Multivariada Análises Fatorial, para extrair os indicadores.

6. RESULTADOS E DISCUSSÕES

6.1 ANÁLISE DESCRITIVA DO PERFIL DAS EMPRESAS DE JI-PARANÁ

Os resultados iniciais do trabalho são referentes à caracterização da quantidade e do perfil das empresas de Ji-paraná segundo as fontes oficiais e extraoficiais do país.

De acordo com os dados fornecidos pelo SEBRAE em 2014 Ji-Paraná possuía 5769 empresas. Sendo que o maior número de empresas possui o porte de Microempreendedor individual (2.782), seguida das Microempresas (1.840), Empresas de pequeno porte (535), Empresa de médio porte (89) e Empresa de grande porte (19). De 504 empresas que não foi informado o porte, veja na tabela 5.

Tabela 5 - Quantidade de empresas ativas em Ji-Paraná, segundo seu porte econômico, de acordo com o SEBRAE.

Porte	Quantidade total
Porte não informado	504
Microempresa	1840
Microempreendedor individual	2782
Empresa de pequeno porte	535
Empresa de médio porte	89
Empresa de grande porte	19
Total	5769

Fonte: SEBRAE, 2014.

Em contra partida, conforme dados disponibilizados pelo IBGE em 2015 Ji-Paraná possui 3.118 empresas ativas. Sendo a segunda maior cidade em quantidade de empresas seguindo a população que também é a segunda maior do estado de Rondônia, observe na tabela 6.

Tabela 6 - Comparativo da proporção de Empresas ativas vs. População, no estado de Rondônia em 2015 de acordo com o IBGE.

Municípios	População	Empresas
Porto Velho	519.436	9627
Ji-Paraná	132.667	3118
Ariquemes	107.345	2397
Vilhena	95.630	2542
Cacoal	88.507	1986

Fonte: IBGE, 2015.

Entretanto as estatísticas mais atuais, de 2015 até em tempo real 2017, referente a quantidade de empresas em Ji-Paraná foram encontradas e disponibilizada pelo site O Empresômetro – Inteligência de mercado, que nasceu de um projeto sócio tecnológico do IBPT – Instituto Brasileiro de Planejamento e Tributação, que vem desde 2006 construindo projetos com base na arrecadação tributária nacional.

O (EMPRESOMETRO, 2017) afirma que em Ji-paraná estão abertas 11.898 empresas, o que representa 8,54% das empresas de Rondônia. Que possui um percentual de crescimento de 1,69% referente o ano de 2016. Concluindo que apesar do momento econômico delicado do país a taxa de empreendedorismo no município cresceu ao invés de retroceder ou estagnar, veja na tabela 7.

Tabela 7 - Número de empresas abertas em Ji-Paraná nos últimos três anos (2015, 2016 e 2017).

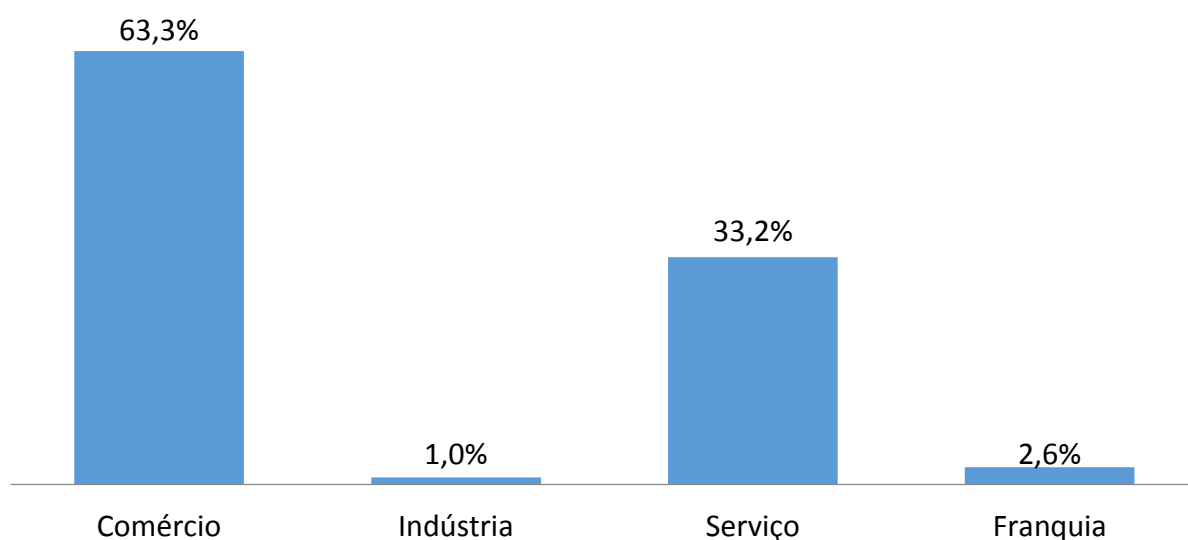
	Ano	Empresas em Ji-Paraná
Nº de empresas	2015	10.600
	2016	11.700
	2017	11.898
Crescimento	2015 a 2016	10,38%
	2016 a 2017	1,69%
% que Ji-Paraná representa no total do estado		8,54%

Fonte: Empresometro, 2017.

É possível observar a seguir a caracterização do perfil das 196 empresas de Ji-paraná que participaram do presente estudo. O seguimento de atuação da maioria das empresas de Ji-Paraná são no setor de Comércio com 63,3% (Lojas de confecções, utensílios domesticos,

alimentício...) e com 33,2% do seguimento de prestação de serviço (manutenção de carro, posto de gasolina, estética, academia) (Figura 1).

Figura 1 - Setor de atuação das 196 empresas de Ji-Paraná entrevistadas por meio de pesquisa exploratória em 2017.



Fonte: Haverroth, 2017.

Dentre as 196 empresas entrevistadas 122 informaram sua renda anual. Onde as empresas possuem uma média de R\$ 250.000,00 em sua renda anual. O que equivale a aproximadamente R\$ 20.800,00 por mês.

Em relação ao porte da empresa na tabela 8 observamos os dois critérios comumente utilizados para definir o porte das empresas, utilizados pelo SEBRAE, no entanto tais critérios não possui fundamentação legal, para fins legais, vale o previsto na legislação do Simples (Lei 123 de 15 de dezembro de 2006).

Verifica-se também na tabela 8 a quantidade de empresas entrevistadas que se encaixam em cada característica dos portes das empresas, verifica-se que tanto no critério de renda bruta como número de empregados as empresas de Ji-Paraná/RO, entrevistadas possuem um porte de Microempresas e Empresas de pequeno porte, em ambos os critérios.

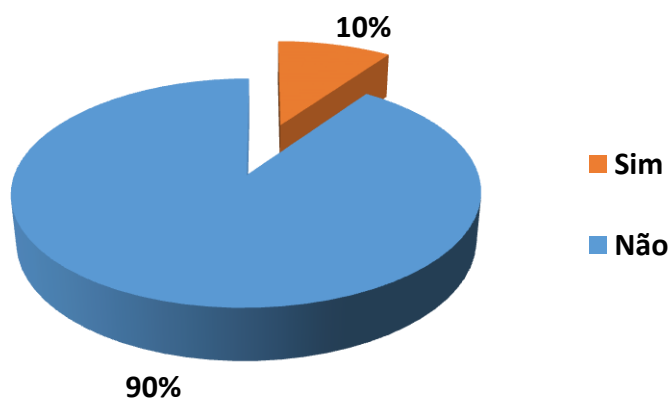
Tabela 8 - Empresas entrevistadas em Ji-Paraná/2017, por meio de uma pesquisa exploratória de acordo com os critérios de classificação do porte econômico, utilizado pelo SEBRAE (tamanho amostral = 196).

Porte da empresa	
1º critério Quanto à Receita Bruta Anual	Nº de empresas
Microempreendedor Individual - Até R\$ 60.000,00	30
Microempresa - Até R\$ 360.000,00	42
Empresa de Pequeno Porte - EPP - De R\$ 360.000,01 até R\$3.600.000,00	50
Não informaram	74
Total	196
2º critério Quanto ao número de Empregados	Nº de Funcionários
Micro: até 9 empregados	149
Pequena: de 10 a 49 empregados	39
Média: de 50 a 99 empregados	1
Grande: mais de 100 empregados	1
Não souberam informar	6
Total	196

Fonte: Haverroth, 2017.

É possível observar na figura 2 que o trabalho também contribui para identificar que a maioria das empresas entrevistas reconhece a importância da utilização de análise dados, no entanto das 196 empresas entrevistadas, apenas 19 já investiram em algum momento em serviços de análises, especificadamente de dados externos (pesquisa de mercado e mídia).

Figura 2 - Percentual das 196 empresas entrevistadas que investem em análise de dados em Ji-Paraná/RO.



Fonte: Haverroth, 2017.

Dessas 19 empresas que disseram investir em análises de dados, verificamos uma desvalorização do profissional da área, pois o capital de investimento dessas empresas neste serviço de análise de dados foi em média de R\$500,00.

Ao serem questionadas se haviam interesse em investir nessas ferramentas para extrair informações úteis para suas tomadas de decisões. As empresas responderam quase que unânimes “não”, justificando suas respostas com os argumentos de que “não viam necessidade”, “não tinham interesse”, “é muito caro” e “o momento financeiro do país não está contribuindo para esse tipo de investimento”.

Entretanto foi possível verificar que as empresas de Ji-Paraná armazenam em média 15 anos de dados internos referentes a cadastro dos clientes, fluxo de caixa, compra e venda de mercadoria, ou seja, as empresas não utilizam os dados armazenados para tomada de decisões, mas os armazenam.

6.2 APLICAÇÃO DA ANÁLISE FATORIAL NOS DADOS

6.2.1 VERIFICAÇÃO DA EFICIÊNCIA DO QUESTIONÁRIO

Antes de iniciar a análise verificou-se que a amostra apresentou 7,84 (196/25) casos por variáveis, atendendo as recomendações de (HAIR, BLACK e BABIN, 2009). E através do coeficiente alfa de Cronbach estimou-se um $\alpha = 0,74$ (196 respostas para 25 variáveis) de confiabilidade do instrumento de registro utilizado, através da correlação média entre as respostas dos entrevistados e as perguntas do instrumento de registro que utilizou a mesma escala de medição.

6.2.2 NORMALIDADE

A princípio para aplicar a Análise Fatorial verifica-se se dados seguem uma distribuição de normalidade, para adiante definir qual o método de estimação dos fatores deve ser utilizado. Veja na tabela 9 o resultado do teste de normalidade de Shapiro-Wilk univariado e multivariado, obtidos através do Software R.

Tabela 9 - Resultado do teste de normalidade de Shapiro-Wilk da pesquisa com empresários, realizada em Ji-Paraná em 196 empresas.

Teste Univariado					
Variáveis	Estatística do Teste - W	p-valor	Variáveis	Estatística do Teste - W	p-valor
Var1	0.7536	0	Var14	0.6325	0
Var2	0.5178	0	Var15	0.7509	0
Var3	0.4788	0	Var16	0.6884	0
Var4	0.7408	0	Var17	0.6228	0
Var5	0.7849	0	Var18	0.7613	0
Var6	0.6960	0	Var19	0.8180	0
Var7	0.6030	0	Var20	0.7885	0
Var8	0.7538	0	Var21	0.8086	0
Var9	0.6696	0	Var22	0.7109	0
Var10	0.7182	0	Var23	0.8292	0
Var11	0.6651	0	Var24	0.3993	0
Var12	0.6148	0	Var25	0.7644	0
Var13	0.5779	0			
Teste Multivariado W = 0.044749, p-valor 0,05 < 2.26					

Fonte: Haverroth, 2017.

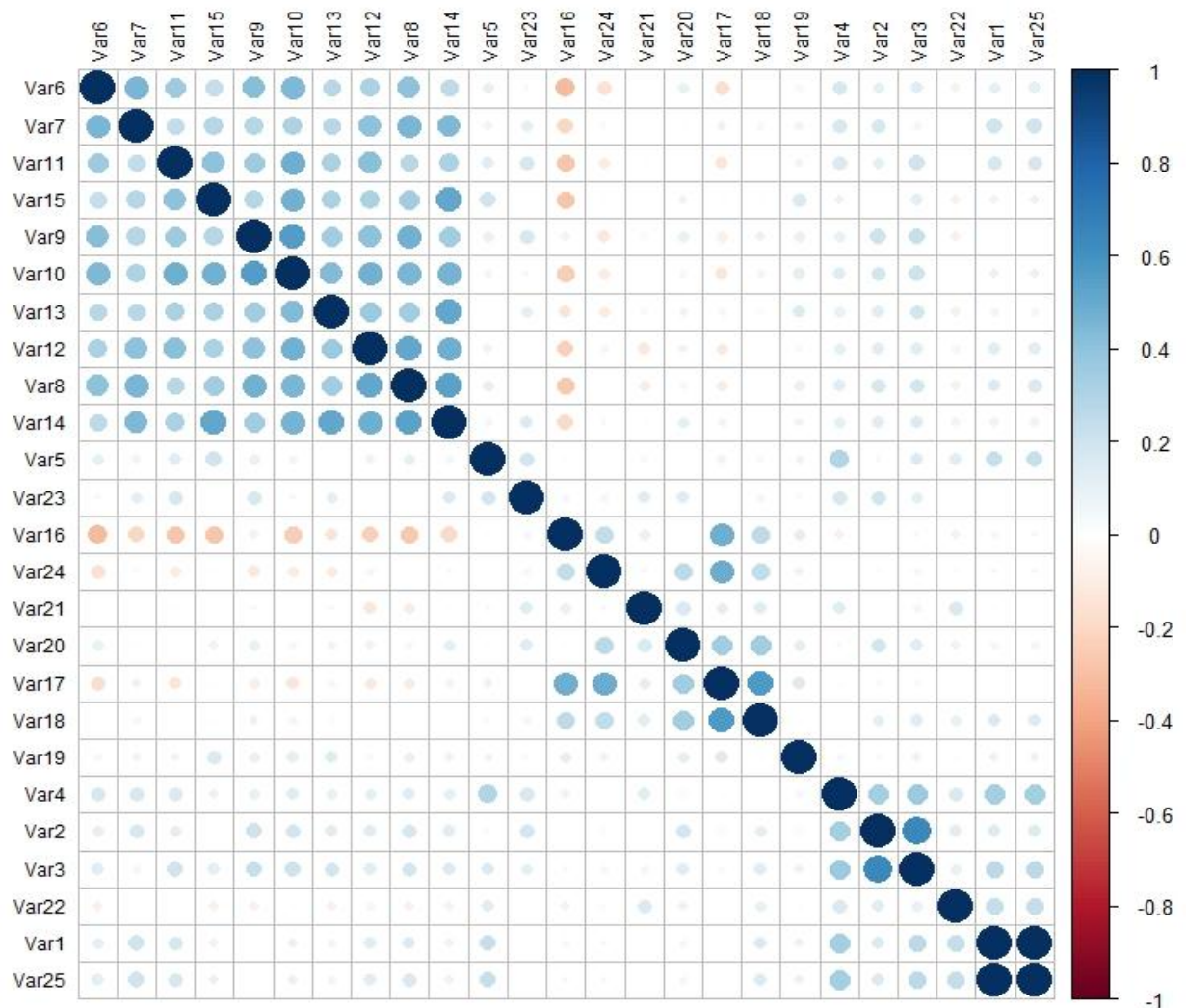
Os dados não possuem uma distribuição normal, nível de significância de 5%, tanto para as variáveis individuais como todas as variáveis juntas, o que já era esperado uma vez que os dados são qualitativos ordinais (de 5 alternativas) o que consequentemente levaria a não normalização dos dados.

Sendo assim utilizar ainda a Análise Fatorial mesmo sem a normalidade dos dados, através da matriz fatorial estimada pelo método de componentes principais, declarado mais adiante, e não mais através do método de máxima verossimilhança que não dispensa a normalidade multivariada. Conforme (MINGOTI, 2013).

6.2.3 TESTANDO A CORRELAÇÃO

Após testar a normalidade dos dados, verificou-se a correlação das variáveis, pressuposto fundamental e indispensável para realizar a AF. Foi utilizada a matriz de correlação a fim de verificar se esta não é uma a matriz identidade, o que significa que as variáveis estão correlacionadas. Outra forma de verificar se o conjunto de dados é adequado é o teste KMO e MSA (por variável).

Figura 3 - Matriz de Correlação das 25 variáveis



Fonte: Haverroth, 2017.

Para a matriz apresentada na figura 3, verificou-se que esta não é uma matriz identidade através do Teste de esfericidade de Bartlett. Este teste produziu um p-valor de 0 a partir da estatística Qui-Quadrado 2538,09 com 300 graus de liberdade. Isso representa que a matriz de correlação não é uma matriz identidade ao nível de significância de 5%, ou seja, há uma associação entre as variáveis.

Para a mesma matriz de correlação foi aplicado também o teste KMO para adequação da amostra considerando as 25 variáveis da tabela 4 que resultou no valor de 0,73 com classificação “Boa” para efetivar a Análise Fatorial.

Outra medida verificada foi a medida de adequação de amostras MSA que para cada variável resultou nos seguintes valores:

Tabela 10 - Medida de adequação de cada variável (I)

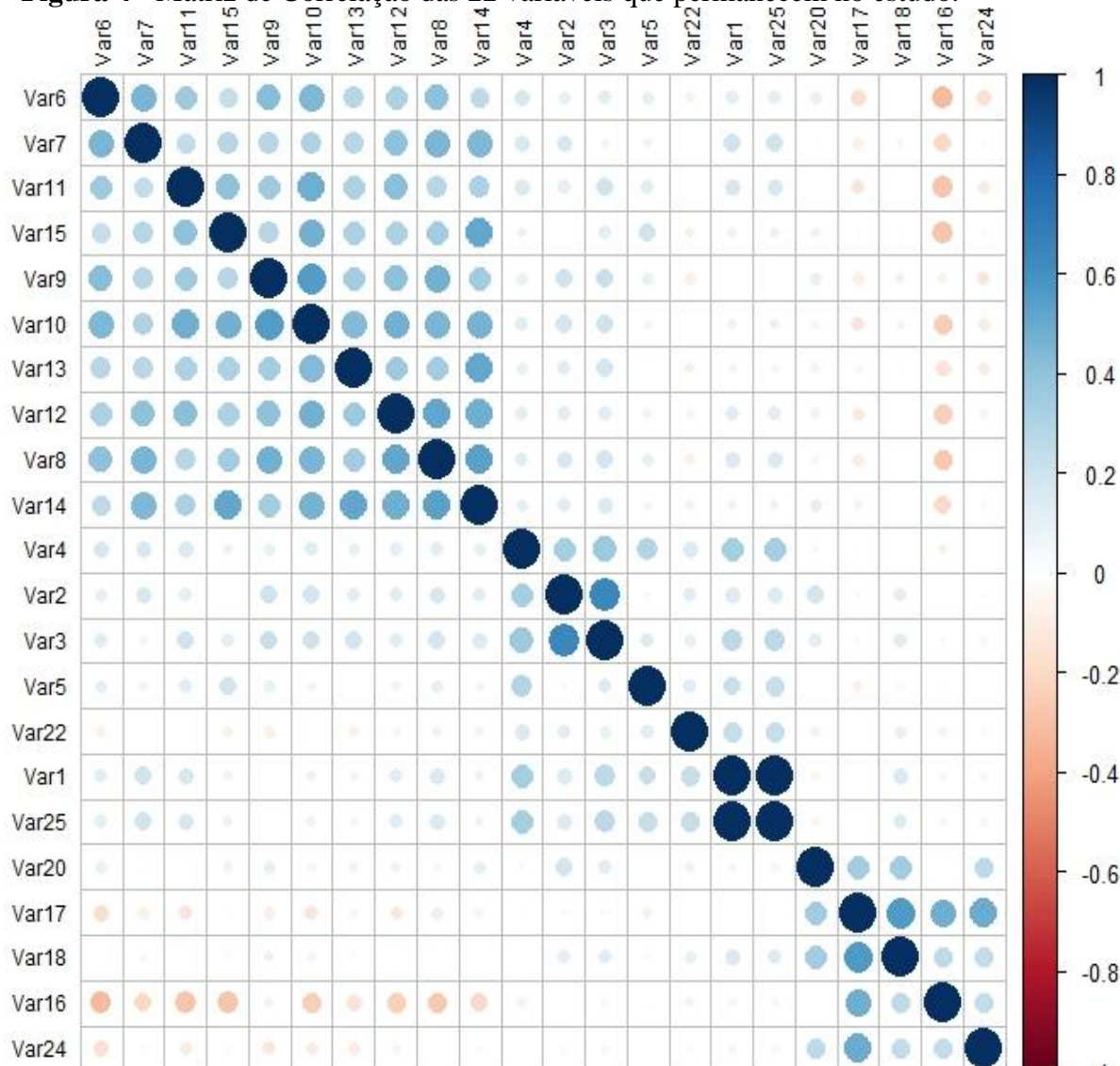
Variáveis	MSA	Variáveis	MSA
Var 1	0.58	Var 14	0.83
Var 2	0.59	Var 15	0.78
Var 3	0.68	Var 16	0.69
Var 4	0.78	Var 17	0.62
Var 5	0.59	Var 18	0.64
Var 6	0.82	Var 19	0.47
Var 7	0.81	Var 20	0.57
Var 8	0.85	Var 21	0.47
Var 9	0.83	Var 22	0.68
Var 10	0.87	Var 23	0.50
Var 11	0.86	Var 24	0.64
Var 12	0.89	Var 25	0.58
Var 13	0.89		

Fonte: Haverroth, 2017.

Todas as medidas MSA estão apresentadas na tabela 10, para cada variável que não se encontrou no domínio aceitável de ($0,5 < \text{MSA} < 0,75$) houve a necessidade de retirá-las da análise para realização da Análise Fatorial, neste caso foram três variáveis (Var 19, Var 21, Var 23). Realizou-se os mesmos procedimentos sem as três variáveis, teste amostral, alfa de Cronbach, o teste da matriz de correlação, KMO e MSA. Porém não será necessário repetir o teste de normalidade, que independente da quantidade de variáveis já teria apresentado normalidade se tivesse.

Para a matriz apresentada na figura 4, o teste de esfericidade de Bartlett produziu um p-valor de 0 a partir da estatística Qui-Quadrado 2443,757 com 231 graus de liberdade. Isso representa que a matriz de correlação não é uma matriz identidade ao nível de significância de 5% havendo uma associação entre as variáveis. A amostra apresentou 8,9 (196/22) casos por variáveis. O coeficiente alfa de Cronbach estimou-se um $\alpha = 0,75$ (196 respostas para 22 variáveis) de confiabilidade do instrumento de registro aplicado.

Figura 4 - Matriz de Correlação das 22 variáveis que permanecem no estudo.



Fonte: Haverroth, 2017.

O teste KMO para adequação da amostra considerando as 22 variáveis resultou em 0.75, com classificação “Boa”, medida aceitável para aplicação da Análise Fatorial e a medida MSA para cada variável foram as seguintes:

Todas as medidas MSA apresentadas na tabela 11, a seguir, que se encontram no domínio aceitável de $(0, 5 < MSA < 0, 75)$, portanto com base na MSA não há necessidade de eliminar mais nenhuma variável para a aplicação análise fatorial, até o presente momento.

Após verificação do teste de esfericidade de Barlett e das medidas de adequação da amostra KMO e MSA, o conjunto de dados é aceitável para continuar Análise Fatorial.

Tabela 11 - Medida de adequação de cada variável (II)

Variáveis	MSA	Variáveis	MSA
Var 1	0.58	Var 12	0.91
Var 2	0.61	Var 13	0.89
Var 3	0.68	Var 14	0.84
Var 4	0.81	Var 15	0.79
Var 5	0.62	Var 16	0.69
Var 6	0.82	Var 17	0.61
Var 7	0.81	Var 18	0.65
Var 8	0.86	Var 20	0.59
Var 9	0.83	Var 22	0.72
Var 10	0.86	Var 24	0.71
Var 11	0.89	Var 25	0.58

Fonte: Haverroth, 2017.

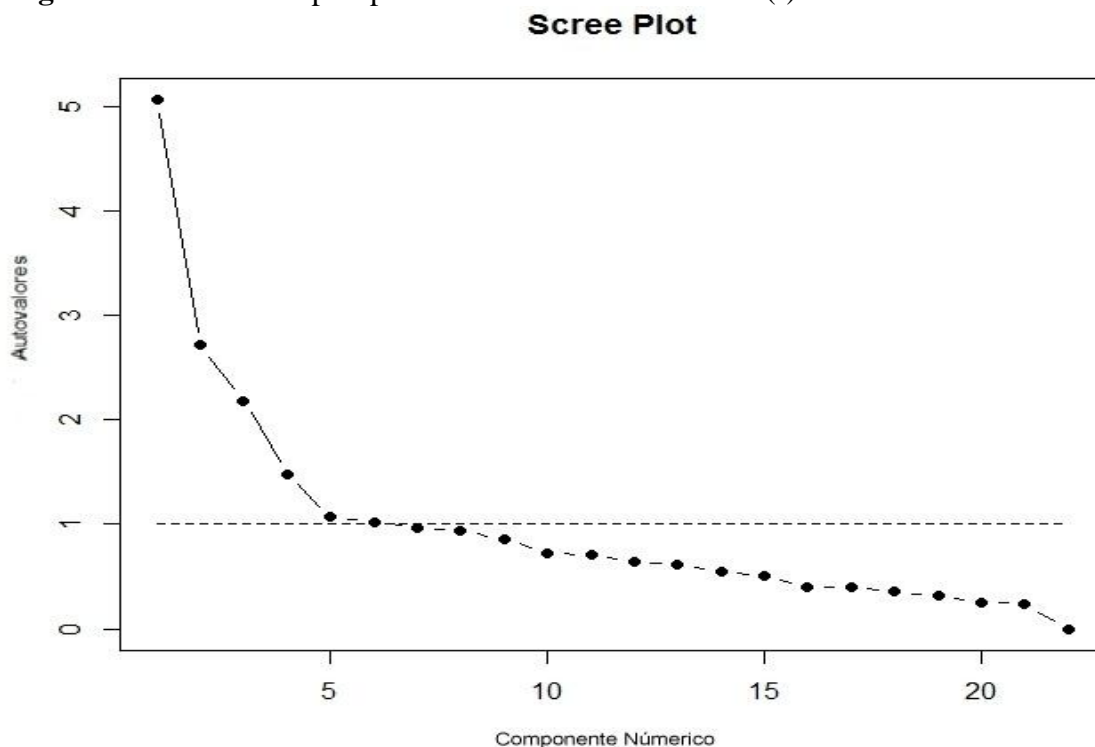
6.2.4 DEFININDO O NÚMERO DE FATORES

A partir da matriz de correlação dada foram encontrados os autovalores e autovetores que serão os responsáveis para determinar o número de componentes a conservar, assim como para calcular as cargas fatoriais para a formulação do modelo ortogonal fatorial, utilizando o método de componentes principais.

Como os autovalores foram extraídos a partir da matriz de correlação e considerando o método da raiz latente, então retemos 4 fatores, pois temos 4 autovalores maiores que 1.

O teste Scree representado na figura 5, afirma que quatro fatores podem ser definidos. A linha tracejada ilustra o critério da raiz latente mostrando o número de componentes a conservar. Desse modo os quatros fatores retidos explicam 92% da variância total.

Figura 5 - Teste Scree-plot para definir número de fatores (I).



Fonte: Haverroth, 2017.

Para a extração dos fatores foi utilizado o método de componentes principais, pois o conjunto de dados não apresentou normalidade e por este método ser mais apropriado quando o interesse é a redução de dados.

As cargas fatoriais juntamente com a comunalidade de cada variável são apresentadas na tabela 12. É possível observar que as cargas fatoriais das variáveis Var4, Var5, Var6, Var7, Var11, Var13, Var15, Var16, Var20, Var22 não possui cargas significativas, apresentaram comunalidade abaixo dos níveis de explicação aceitáveis recomendados por (HAIR, BLACK e BABIN, 2009).

Sendo assim foi necessário serem retiradas da matriz mais 10 variáveis, fazendo com que todo o processo, realizado até aqui, seja refeito. A variável Var24 apresentar comunalidade de 0,49, mesmo assim irá permanecer na nova análise para verificar se ao retirar as outras 10 variáveis sua carga aumenta, uma vez que esta bem próxima do valor definido como critério.

Tabela 12 - Matriz Fatorial não rotacionada (I).

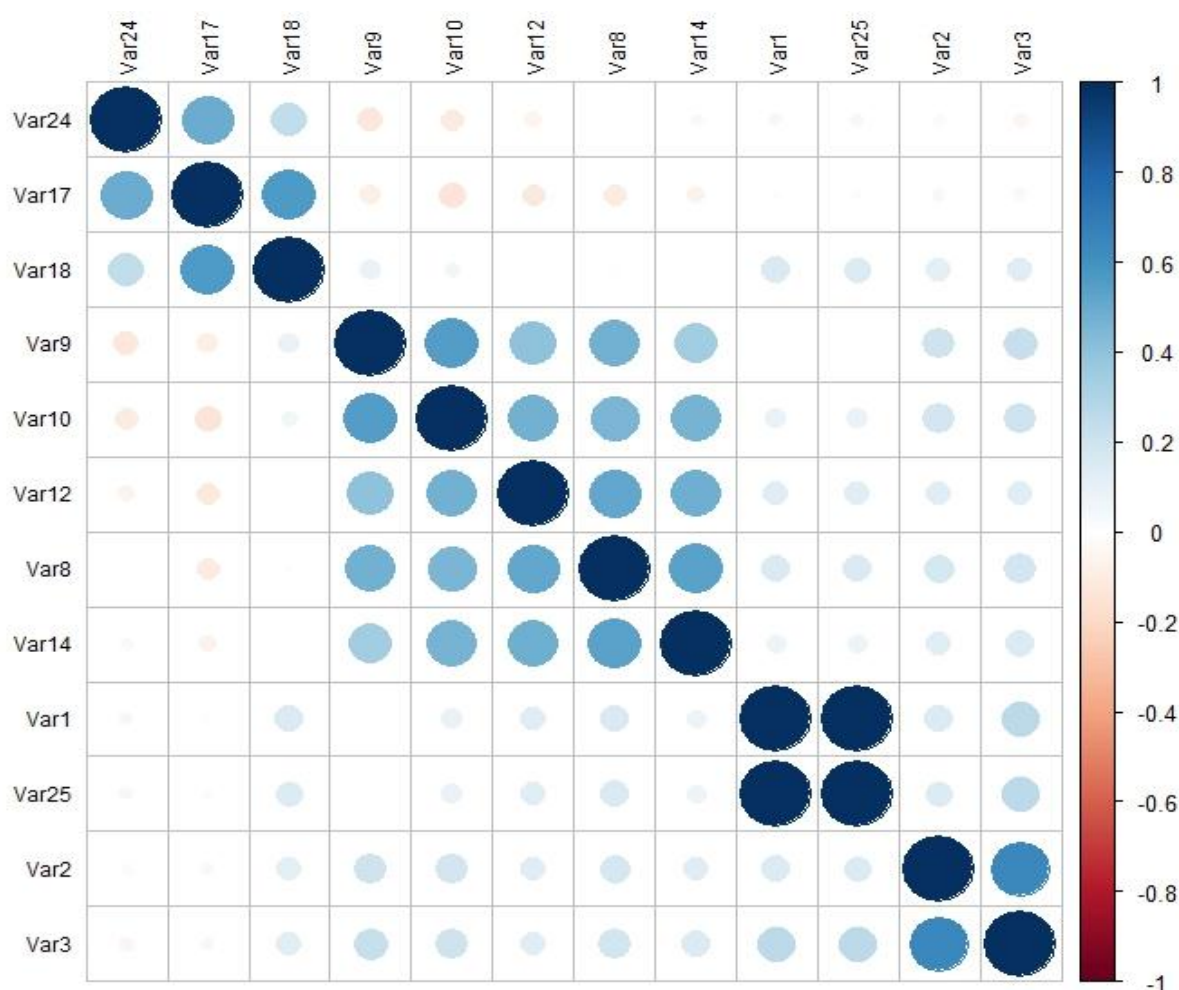
Variáveis	Fatores				Comunalidade
	1	2	3	4	
Var1	0.32	0.67	-0.46	0.36	0.89
Var2	0.33	0.40	0.00	-0.70	0.76
Var3	0.39	0.45	-0.09	-0.61	0.74
Var4	0.34	0.41	-0.32	-0.23	0.44
Var5	0.22	0.26	-0.24	0.11	0.18
Var6	0.61	-0.12	-0.01	0.00	0.39
Var7	0.60	0.00	0.03	0.20	0.41
Var8	0.71	-0.08	0.13	0.10	0.54
Var9	0.63	-0.11	0.23	-0.18	0.50
Var10	0.74	-0.13	0.14	-0.04	0.59
Var11	0.62	-0.05	-0.05	0.05	0.40
Var12	0.68	-0.11	0.13	0.12	0.50
Var13	0.58	-0.13	0.18	-0.02	0.39
Var14	0.69	-0.13	0.22	0.14	0.56
Var15	0.59	-0.12	0.15	0.23	0.44
Var16	-0.39	0.43	0.28	0.04	0.42
Var17	-0.21	0.51	0.67	0.11	0.77
Var18	0.02	0.53	0.51	0.07	0.55
Var20	0.08	0.27	0.55	-0.14	0.41
Var22	-0.01	0.40	-0.22	-0.02	0.21
Var24	-0.15	0.39	0.48	0.29	0.49
Var25	0.32	0.67	-0.47	0.36	0.89

Fonte: Haverroth, 2017.

Refazendo os testes e procedimentos novamente sem as dez variáveis (Var4, Var5, Var6, Var7, Var11, Var13, Var15, Var16, Var20, Var22), observa-se os seguintes resultados.

A amostra apresentou 16,33 (196/12) casos por variáveis. O coeficiente alfa de Cronbach estimou-se um $\alpha = 0,68$ (196 respostas para 12 variáveis) de confiabilidade do instrumento de registro aplicado. Apesar de ter diminuído o valor do α , ainda temos um resultado significativo.

Figura 6 - Matriz de Correlação das 12 variáveis que permanecem no estudo.



Fonte: Haverroth, 2017.

Teste de esfericidade de Bartlett produziu um p-valor de 0 a partir da estatística Qui-Quadrado 1792,182 com 66 graus de liberdade ao nível de significância de 5%, o que significa que há uma associação entre as variáveis.

O teste KMO para adequação da amostra considerando as 12 variáveis resultou em 0.66, o que representa uma medida aceitável, segundo a tabela de critério de decisão do teste para aplicação da Análise Fatorial. As novas medidas de MSA para cada variável foram as seguintes:

Tabela 13 - Medida de adequação de cada variável (III).

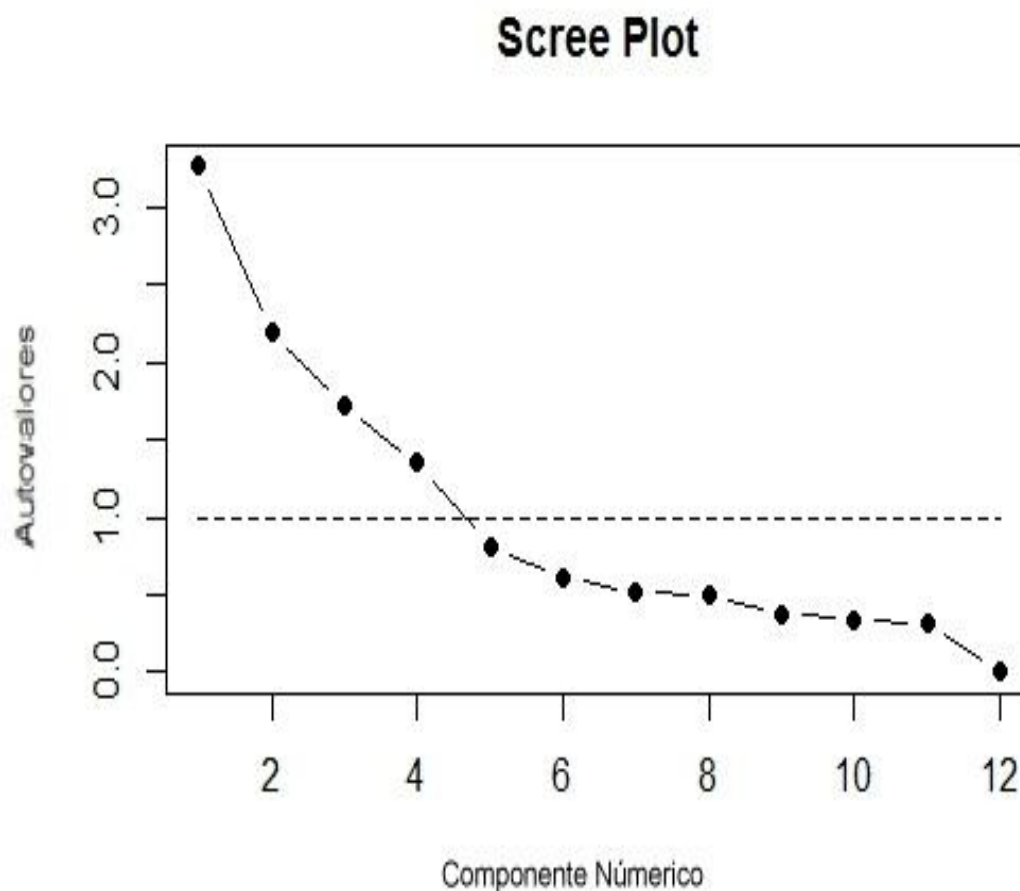
Variáveis	MSA
Var1	0.53
Var2	0.60
Var3	0.64
Var8	0.81
Var9	0.79
Var10	0.82
Var12	0.86
Var14	0.81
Var17	0.55
Var18	0.59
Var24	0.59
Var25	0.53

Fonte: Haverroth, 2017.

Todas as medidas MSA das variáveis apresentadas na tabela 13 atingiram o domínio aceitável de $(0, 5 < \text{MSA} < 0, 75)$, por tanto, não há necessidade de retirar mais variáveis para realizar a Análise Fatorial. Após determinar o número de variáveis a serem mantidos para a análise, foi extraídos os autovalores a partir da matriz de correlação considerando o método da raiz latente, continuando com 4 fatores, pois existem 4 autos valores maiores que 1.

O teste Scree representado na figura 7, afirma que quatro fatores podem ser definidos. A linha tracejada ilustra o critério da raiz latente mostrando o número de componentes a conservar. Desse modo os quatros fatores retidos continuam explicando 92% da variância total.

Figura 7 - Teste Scree-plot para definir número de fatores (II).



Fonte: Haverroth, 2017.

6.2.5 EXTRAÇÃO DOS FATORES PELO MÉTODO DE COMPONENTE PRINCIPAL

Na extração dos fatores foi utilizado o método de componentes principais, seguem na tabela 14 as cargas fatoriais juntamente com a comunalidade de cada variável.

O percentual de variância na solução não rotacionada dos fatores explicam, respectivamente, 27%, 18%, 14% e 11 da variância. Ao observar que todas as variáveis da matriz não rotacionada obteve carga significativa, realiza-se a rotação da matriz fatorial.

Tabela 14 - Matriz Fatorial não rotacionada II.

Variáveis	Fatores				Comunalidade	Variância Específica
	1	2	3	4		
Var1	0.41	0.69	-0.51	0.28	0.99	0.0093
Var2	0.45	0.26	0.03	-0.73	0.80	0.1980
Var3	0.51	0.34	-0.06	-0.65	0.80	0.1966
Var8	0.74	-0.18	0.14	0.21	0.64	0.3647
Var9	0.66	-0.27	0.23	-0.07	0.57	0.4322
Var10	0.72	-0.25	0.15	0.05	0.60	0.3953
Var12	0.69	-0.22	0.11	0.25	0.60	0.4040
Var14	0.66	-0.23	0.18	0.22	0.57	0.4300
Var17	-0.15	0.54	0.70	0.05	0.80	0.1952
Var18	0.10	0.55	0.54	0.03	0.61	0.3921
Var24	-0.12	0.44	0.53	0.26	0.55	0.4513
Var25	0.41	0.69	-0.51	0.29	0.99	0.0082
Autovalor	3.26	2.19	1.72	1.35	-	-
Proporção da Variância	0.27	0.18	0.14	0.11	-	-
Variância Acumulada	0.27	0.45	0.60	0.71	-	-
Proporção Explicada	0.38	0.26	0.20	0.16	-	-
Proporção Acumulada	0.38	0.64	0.84	1.00	-	-

Fonte: Haverroth, 2017.

6.2.6 ROTAÇÃO ORTOGONAL DA MATRIZ FATORIAL - VARIMAX

Foram utilizados os critérios de rotação ortogonal Varimax para analisar se existe alguma diferença dos comportamentos das variáveis nos distintos fatores. Os resultados das comunalidades foram idênticos antes e depois da rotação ortogonal Varimax, que foram os seguintes:

Tabela 15 - Rotação ortogonal da matriz fatorial – Varimax.

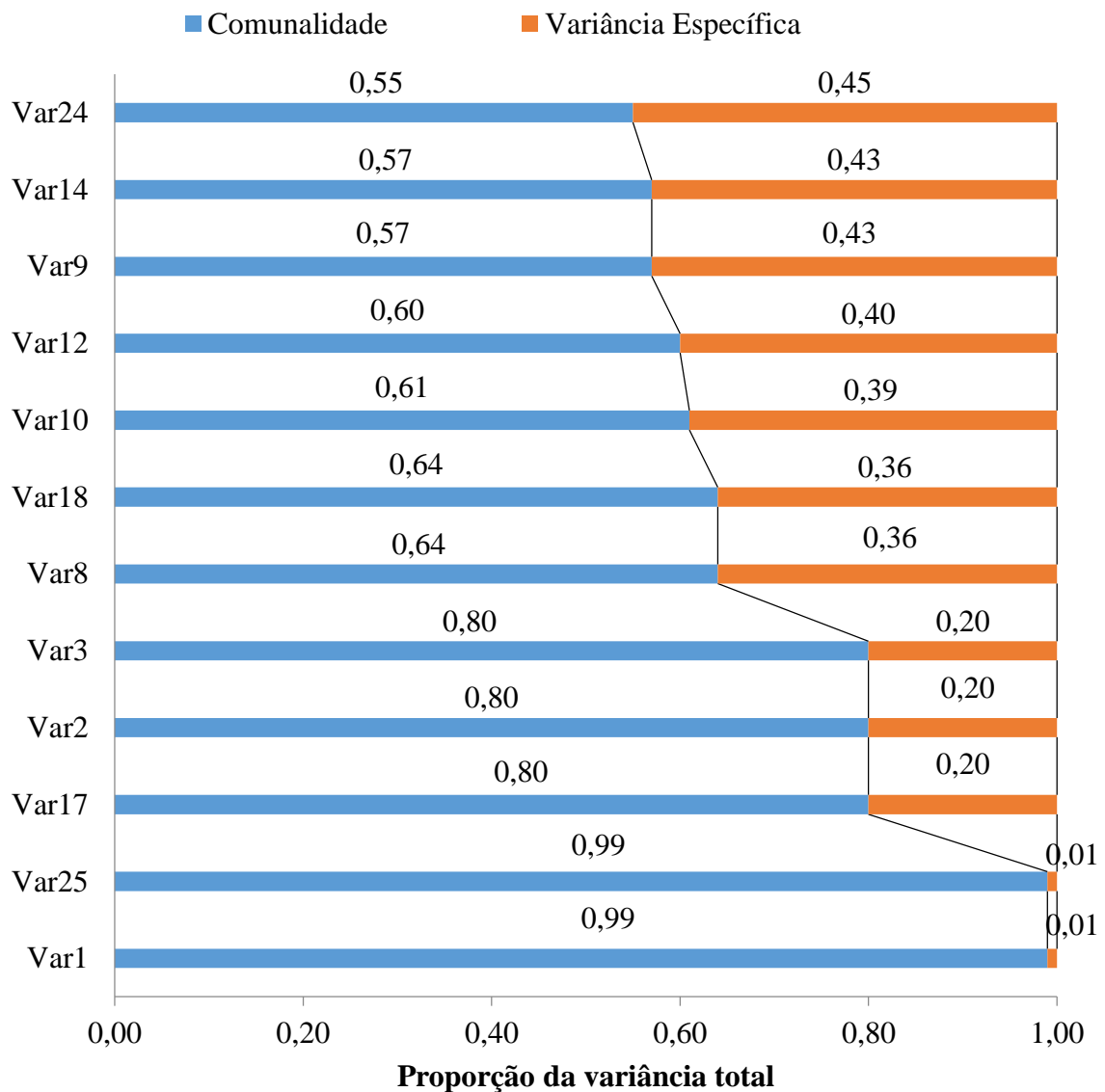
Variáveis	Fatores				Variância	
	1	2	3	4	Comunalidade	Específica
Var1	0.06	0.99	0.04	0.11	0.99	0.0093
Var2	0.12	0.04	0.03	0.89	0.80	0.1980
Var3	0.13	0.19	0.02	0.87	0.80	0.1966
Var8	0.79	0.12	0.00	0.05	0.64	0.3647
Var9	0.70	-0.12	-0.03	0.24	0.57	0.4322
Var10	0.76	0.00	-0.06	0.16	0.60	0.3953
Var12	0.76	0.11	-0.04	-0.02	0.60	0.4040
Var14	0.75	0.04	0.00	-0.01	0.57	0.4300
Var17	-0.11	-0.06	0.89	0.05	0.80	0.1952
Var18	0.05	0.10	0.75	0.17	0.61	0.3921
Var24	-0.04	0.04	0.72	-0.15	0.55	0.4513
Var25	0.06	0.99	0.04	0.11	0.99	0.0082
Autovalor	2.90	2.04	1.89	1.70	-	-
Proporção da Variância	0.24	0.17	0.16	0.14	-	-
Variância Acumulada	0.24	0.41	0.57	0.71	-	-
Proporção Explicada	0.34	0.24	0.22	0.20	-	-
Proporção Acumulada	0.34	0.58	0.80	1.00	-	-

Fonte: Haverroth, 2017.

Observa-se na tabela 15 que o percentual de variância na solução rotacionada dos fatores explicam, respectivamente, 24%, 17%, 16% e 14% da variância. Nota-se que apesar de uma pequena alteração o percentual de variância, foi redistribuído de forma equilibrada com as rotações Varimax. Onde os valores das comunalidades se mantiveram iguais antes e depois da rotação.

Veja na figura 8 a variância total, a partir do somatório da variância comum e específica de cada variável.

Figura 8 - Visualização da Variância Total.

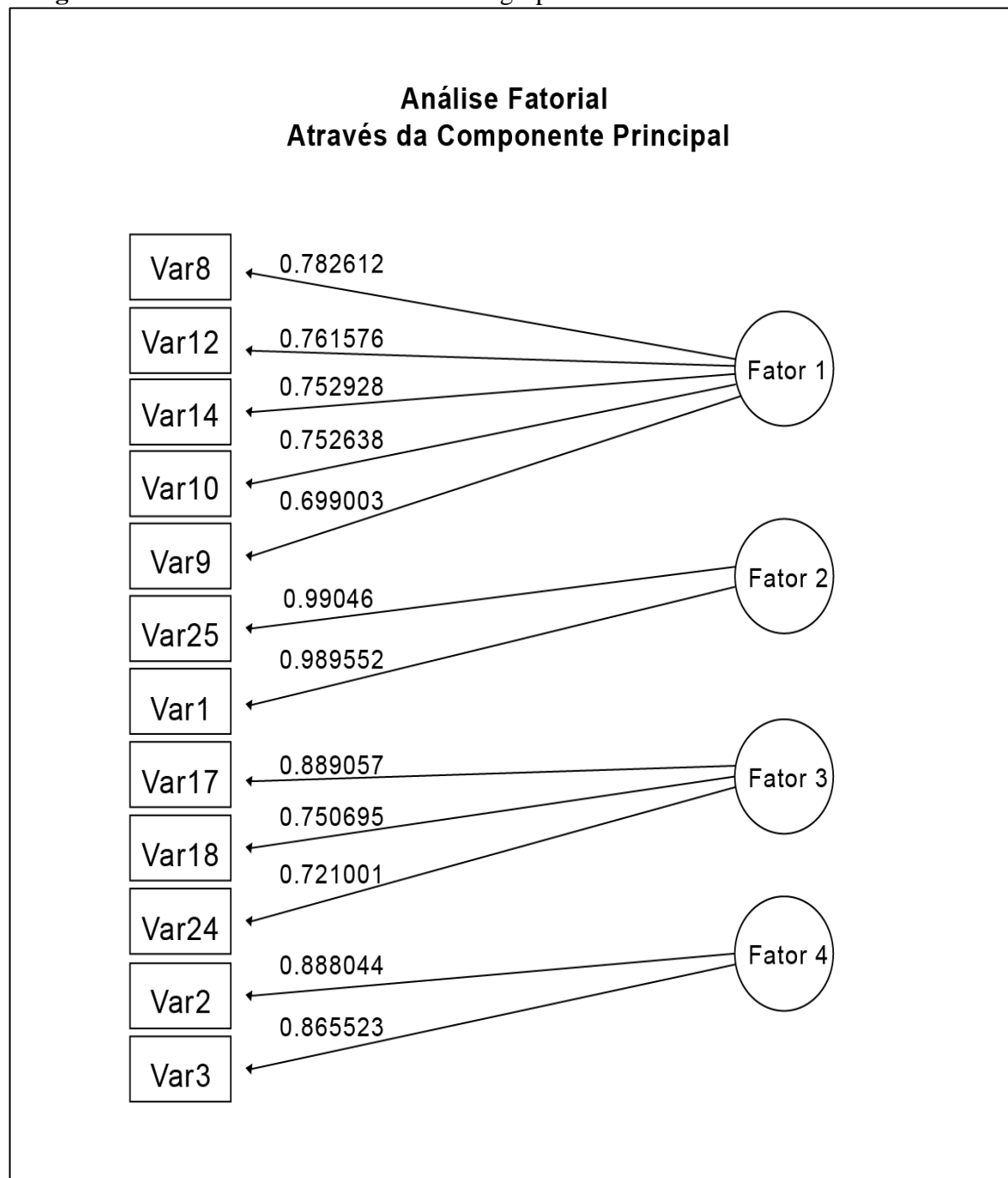


Fonte: Haverroth, 2017.

6.2.7 INTERPRETAÇÃO DOS FATORES

Na seção 6.2.6 observou-se que a rotação foi necessária para auxiliar na interpretação das matrizes fatorial, de forma a torná-la simplificada. Para classificar quais cargas são significativas na matriz fatorial, utilizamos o critério citado por (HAIR, BLACK e BABIN, 2009). No presente estudo, cujo tamanho amostral é igual a 196 indivíduos, considera cargas significativas aquelas maiores que 0,50. Na figura 9 é possível observar a distribuição das variáveis em seus respectivos fatores.

Figura 9 - Variáveis latentes extraídas do agrupamento das variáveis através da AF.



Fonte: Haverroth, 2017.

A partir dos fatores acima, definiu-se o fator 1 como "Benefícios", o fator 2 como "Inovação", o fator 3 como "Dificuldades" e o fator 4 como "Organização".

Tabela 16 - Interpretação dos fatores a partir do critério de rotação Varimax.

ANÁLISE FATORIAL		CARGA FATORIAL
FATOR1	BENEFÍCIOS	
Var8	Auxiliar na tomada de decisões	0.78
Var12	Conhecer falhas internas da empresa	0.76
Var10	Criar promoções	0.75
Var14	Melhorar gestão	0.75
Var9	Ajuda na hora de comprar mais mercadoria	0.70
FATOR2	INOVAÇÃO	
Var1	A empresa não tem registro cadastral dos clientes	0.98
Var25	Não gosto de tecnologia	0.98
FATOR3	DIFICULDADES	
Var17	Na pesquisa não terá nada que eu já não saiba sobre minha empresa e meus clientes	0.88
Var18	Não consigo colocar em prática as informações extraídas da pesquisa.	0.75
Var24	Não gosto de ser avaliado	0.72
FATOR4	ORGANIZAÇÃO	
Var2	A empresa registra a quantidade e itens de produtos que compra	0.88
Var3	A empresa registra a quantidade e itens de produtos que vende	0.86

Fonte: Haverroth, 2017.

Na tabela 16 observam-se os quatro fatores identificados, através da Análise Fatorial com o critério de rotação Varimax, que definem a demanda de contratação e utilização de análise de dados externos e internos por parte das empresas de Ji-Paraná/RO. Foram quatro grupos distintos de 12 variáveis formando os Fatores, Benefício, Inovação, Dificuldades e Organização.

No Fator Benefícios apresenta-se, por assim dizer, os pontos positivos que os empresários acreditam haver ao contratarem serviços de análises de dados internos ou externos. O que será útil quando no futuro forem estabelecidas estratégias para fomentar a cultura de utilização dessa ferramenta nas empresas e assim ampliar o mercado de trabalho para os futuros estatísticos formados pela UNIR de Ji-Paraná.

No Fator Inovação, observa-se uma questão muito interessante que tende a diminuir a demanda de prestação de serviço de análise de dados. Pois alguns dos empresários têm certa dificuldade em aderir os avanços tecnológicos, o que os leva a não armazenar os dados gerados diariamente pelas empresas e a falta de interesse em investir em análises estatísticas.

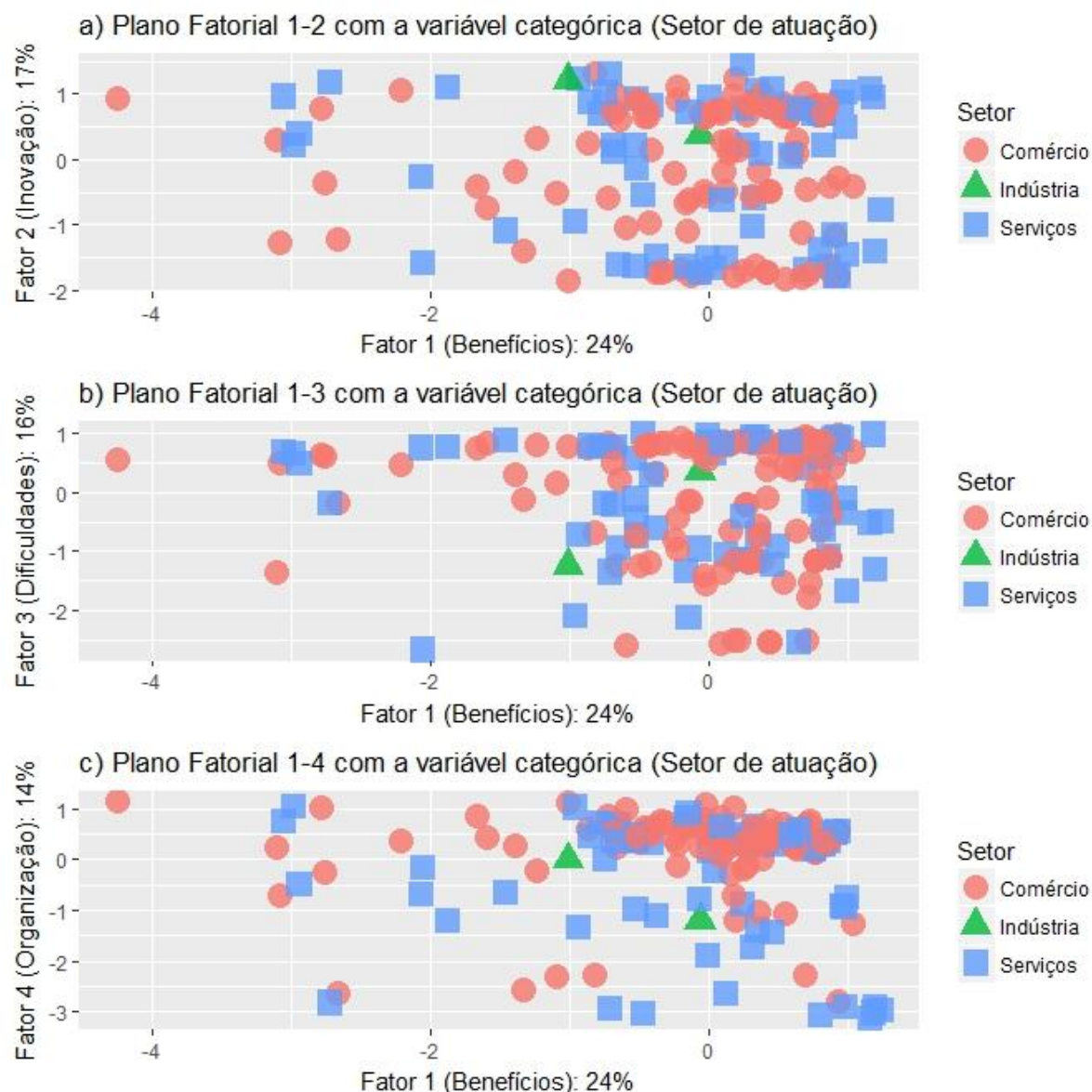
No Fator Dificuldades o próprio nome já diz ao que se refere, este grupo é formado por variáveis que representam as principais ideias que impedem os empresários de contratarem esse tipo de serviço. O interessante é que as três variáveis que compõe este fator não são características propriamente ditas das pesquisas e análises de dados e sim características das crenças culturais dos empresários, bem como sua forma de agir a partir de uma mudança ou avaliação. Entender o significado deste fator é essencial para o profissional estatístico que pretende atuar no seguimento de Análise de Negócios, pois demonstra as dificuldades que se apresentarão em sua carreira.

No Fator Organização apresentam-se variáveis que reforça a importância de se investir em analisar dados. Após a revolução tecnológica dos últimos 20 anos, onde tudo passou a ser automatizado, as empresas agora querem e precisam fazer algo de útil com tantos dados armazenados e neste fator foi possível verificar que as empresas de Ji-Paraná possuem “estoques” de dados a serem explorados. O que vem a ser um incentivo para os futuros profissionais estatísticos, pois ainda que muitas empresas não enxerguem, a necessidade deste profissional já é realidade no município.

Para uma melhor interpretação dos fatores, variáveis latentes encontradas, observam-se os fatores comparados com as variáveis categóricas “Qual setor econômico que a empresa atua” e “A empresa já contratou serviços de análise de dados”.

Referente os fatores que definem a demanda de análise de dados nas empresas de Ji-Paraná, observa-se que ao comparar os fatores 1-2, 1-3 e 1-4 com a variável categórica “**setor de atuação das empresas entrevistadas**”, é possível identificar homogeneidade no comportamento dos dados, o que demonstra um consenso das respostas dos empresários de acordo com seu setor de atuação (Figura 10).

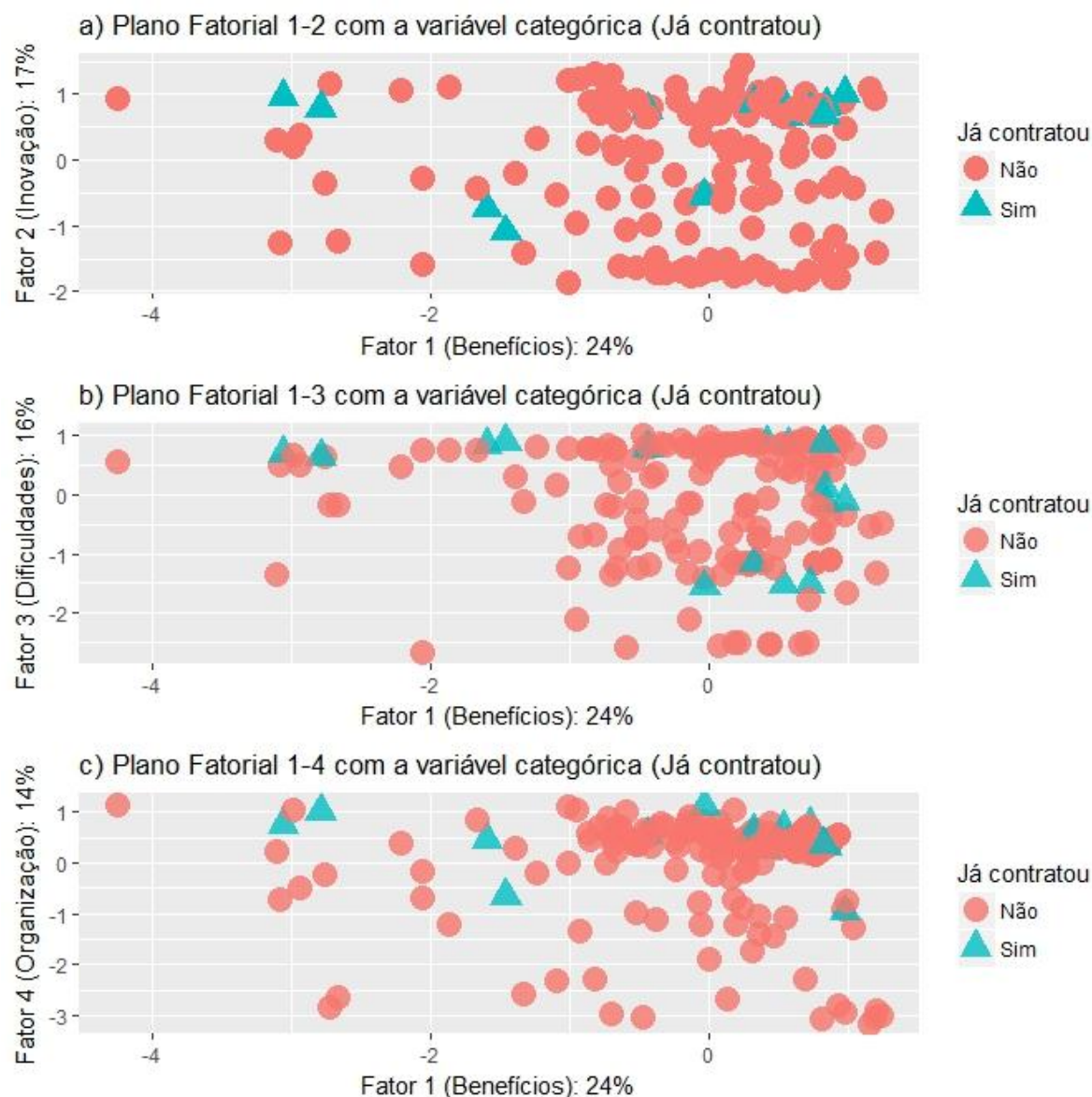
Figura 10 - Comparativo dos gráficos de dispersão dos Scores nos Planos Fatores 1-2, 1-3 e 1-4 com a variável categórica “Setor”.



Fonte: Haverroth.

De acordo com os fatores que definem a demanda de análise de dados nas empresas de Ji-Paraná, observa-se que ao comparar os fatores 1-2, 1-3 e 1-4 com a variável categórica “**já contratou serviços de análise de dados externos ou internos**”, é possível identificar homogeneidade no comportamento dos dados, o que demonstra um consenso das respostas dos empresários de acordo com o fato de investirem em análises dados (Figura 11).

Figura 11 - Comparativo dos gráficos de dispersão dos Scores nos Planos Fatores 1-2, 1-3 e 1-4 com a variável categórica “Já contratou”.



Fonte: Haverroth.

6.2.8 VALIDAÇÃO DOS RESULTADOS

A validação dos resultados utilizou-se por meio da bondade de ajuste. Este procedimento consiste em estimar a matriz residual (14) a partir da diferença entre as correlações observadas e as correlações reproduzidas. A matriz de correlações observadas é dada pela matriz (12) e a matriz de correlações reproduzidas (13) foi estimada conforme a equação (9) apresentada na subseção 4.2.3.

$$\text{Matriz de correlações observadas} = \begin{bmatrix} 1 & 0.15 & 0.29 & 0.17 & -0.04 & 0.06 & 0.14 & 0.08 & -0.02 & 0.15 & 0.05 & 0.99 \\ 0.15 & 1 & 0.79 & 0.14 & 0.28 & 0.22 & 0.07 & 0.08 & 0.05 & 0.18 & -0.11 & 0.14 \\ 0.29 & 0.79 & 1 & 0.17 & 0.27 & 0.23 & 0.10 & 0.10 & 0.03 & 0.19 & -0.11 & 0.28 \\ 0.17 & 0.14 & 0.17 & 1 & 0.55 & 0.60 & 0.61 & 0.59 & -0.09 & 0.05 & -0.03 & 0.17 \\ -0.04 & 0.28 & 0.27 & 0.55 & 1 & 0.57 & 0.52 & 0.52 & -0.08 & 0.04 & -0.09 & -0.04 \\ 0.06 & 0.22 & 0.23 & 0.60 & 0.57 & 1 & 0.57 & 0.57 & -0.13 & 0.01 & -0.10 & 0.06 \\ 0.14 & 0.07 & 0.10 & 0.61 & 0.52 & 0.57 & 1 & 0.58 & -0.12 & 0.01 & -0.05 & 0.14 \\ 0.08 & 0.08 & 0.10 & 0.59 & 0.52 & 0.57 & 0.58 & 1 & -0.08 & 0.04 & -0.03 & 0.08 \\ -0.02 & 0.05 & 0.03 & -0.09 & -0.08 & -0.13 & -0.12 & -0.08 & 1 & 0.66 & 0.63 & -0.02 \\ 0.15 & 0.18 & 0.19 & 0.05 & 0.04 & 0.01 & 0.01 & 0.04 & 0.66 & 1 & 0.52 & 0.14 \\ 0.05 & -0.11 & -0.11 & -0.03 & -0.09 & -0.10 & -0.05 & -0.03 & 0.63 & 0.52 & 1 & 0.05 \\ 0.99 & 0.14 & 0.28 & 0.17 & -0.04 & 0.06 & 0.14 & 0.08 & -0.02 & 0.14 & 0.05 & 1 \end{bmatrix} \quad (12)$$

$$\text{Matriz de correlações reproduzidas} = \begin{bmatrix} 1 & 0.14 & 0.26 & 0.15 & -0.00 & 0.08 & 0.12 & 0.07 & -0.01 & 0.15 & 0.04 & 0.99 \\ 0.14 & 1 & 0.64 & 0.17 & 0.20 & 0.18 & 0.12 & 0.12 & 0.03 & 0.11 & -0.02 & 0.14 \\ 0.27 & 0.64 & 1 & 0.18 & 0.22 & 0.20 & 0.12 & 0.14 & 0.03 & 0.12 & -0.04 & 0.26 \\ 0.15 & 0.17 & 0.18 & 1 & 0.47 & 0.45 & 0.51 & 0.53 & -0.10 & 0.01 & 0.00 & 0.15 \\ -0.00 & 0.20 & 0.22 & 0.47 & 1 & 0.55 & 0.40 & 0.34 & -0.08 & 0.08 & -0.12 & -0.00 \\ 0.08 & 0.18 & 0.20 & 0.45 & 0.55 & 1 & 0.47 & 0.46 & -0.13 & 0.05 & -0.10 & 0.08 \\ 0.12 & 0.12 & 0.12 & 0.51 & 0.40 & 0.47 & 1 & 0.48 & -0.11 & 0.00 & -0.05 & 0.12 \\ 0.07 & 0.12 & 0.14 & 0.53 & 0.34 & 0.46 & 0.48 & 1 & -0.06 & -0.00 & -0.03 & 0.07 \\ -0.01 & 0.03 & 0.03 & -0.10 & -0.08 & -0.13 & -0.11 & -0.06 & 1 & 0.56 & 0.49 & -0.01 \\ 0.15 & 0.11 & 0.12 & 0.01 & 0.08 & 0.05 & 0.00 & -0.00 & 0.56 & 1 & 0.24 & 0.14 \\ 0.04 & 0.02 & -0.04 & -0.00 & -0.12 & -0.10 & -0.05 & -0.03 & 0.49 & 0.24 & 1 & 0.04 \\ 0.99 & 0.14 & 0.26 & 0.15 & -0.00 & 0.08 & 0.12 & 0.07 & -0.01 & 0.14 & 0.04 & 1 \end{bmatrix} \quad (13)$$

$$\text{Por tanto a matriz residual é dada por} = \begin{bmatrix} 0 & 0.00 & -0.02 & -0.02 & 0.05 & 0.02 & -0.02 & -0.01 & 0.01 & 0.00 & -0.02 & 0.01 \\ 0.00 & 0 & -0.14 & 0.03 & -0.08 & -0.04 & 0.05 & 0.04 & -0.09 & -0.07 & 0.09 & 0.00 \\ -0.02 & -0.14 & 0 & 0.02 & -0.06 & -0.03 & 0.02 & 0.05 & 0.00 & -0.06 & 0.06 & -0.02 \\ -0.02 & 0.03 & 0.02 & 0 & -0.07 & -0.15 & -0.09 & -0.06 & -0.01 & -0.05 & 0.04 & -0.02 \\ 0.05 & -0.08 & -0.06 & -0.07 & 0 & -0.02 & -0.12 & -0.18 & -0.00 & 0.04 & -0.03 & -0.05 \\ 0.02 & -0.04 & -0.03 & -0.15 & -0.02 & 0 & -0.11 & -0.10 & -0.00 & 0.03 & -0.01 & 0.02 \\ -0.02 & 0.05 & 0.02 & -0.09 & -0.12 & -0.11 & 0 & -0.10 & 0.01 & -0.01 & -0.00 & -0.02 \\ -0.01 & 0.04 & 0.05 & -0.06 & -0.18 & -0.10 & -0.10 & 0 & 0.03 & -0.05 & -0.00 & -0.01 \\ 0.01 & -0.02 & 0.00 & -0.01 & -0.00 & -0.00 & 0.01 & 0.03 & 0 & -0.10 & -0.14 & 0.01 \\ 0.00 & -0.07 & -0.06 & -0.05 & 0.04 & 0.03 & -0.01 & -0.05 & -0.10 & 0 & -0.28 & -0.00 \\ -0.02 & 0.09 & 0.06 & 0.04 & -0.03 & -0.01 & -0.00 & -0.00 & -0.14 & -0.28 & 0 & -0.01 \\ 0.01 & 0.00 & -0.02 & -0.02 & -0.05 & 0.02 & -0.02 & 0.01 & 0.01 & -0.00 & -0.01 & 0 \end{bmatrix} \quad (14)$$

Podemos observar que a matriz de correlação estimada (13) pelo modelo fatorial se aproxima muito da matriz de correlação dos dados originais (12), com isso podemos observar que o modelo construído para esse tipo de análise foi satisfatório os resultados, por que a matriz de correlação fatorial realmente se aproxima da matriz de correlação estimada. Podendo observar que a matriz residual (14) tem baixo o valor do quadrado médio da raiz dos resíduos (RMSR) de 0.07.

De acordo com a matriz residual dada em (14) temos que aproximadamente 62% dos resíduos são significativos considerando o ponto de corte igual a 0.05 e 23% são significativos ao ponto de corte de 0.10. Observamos que essa porcentagem é consideravelmente alta sendo 85% abaixo de 0.10, isso pode ser explicado pela quantidade de variância explicada pela solução fatorial em 4 fatores ser igual a 71%.

7. CONCLUSÕES

O presente trabalho teve como propósito principal identificar os fatores que influenciam na demanda, contratação e utilização de análise de dados externos e internos por parte das empresas de Ji-Paraná/RO. Para isso utilizou-se a Análise Fatorial Exploratória estimada pela componente principal e com o critério de rotação Varimax. Através disso, identificou-se a formação de quatro grupos distintos de variáveis com estrutura composta pelas dimensões/fatores (Benefícios, Dificuldades, Inovação e Organização) de um total de 12 variáveis.

Em relação aos objetivos específicos, o trabalho também contribui para identificar que a maioria das empresas entrevistadas reconhece a importância da utilização das análises de dados, porém foram unânimes em dizer que não viam necessidade ou não tinham interesse em investir nesse tipo de serviço. Evidenciando que o mercado ainda não está absorvendo os profissionais que executam esse trabalho, como os estatísticos recém-formados no município. Essa falta de interesse do mercado local é um dos motivos que impedem a contratação desses serviços nas empresas. Por consequência pode ser um dos motivos que desmotiva os alunos do curso de graduação em estatística concluírem curso.

Das 19 empresas que disseram investir em análises de dados verificamos uma desvalorização do profissional da área, pois o capital de investimento dessas empresas foi em média de R\$500,00.

Não foi possível verificar uma relação do faturamento anual da empresa e a efetiva contratação de pesquisas, pois o número de empresas entrevistadas que investem nesse seguimento não foram suficientes para se chegar a essa informação.

8. PROPOSIÇÕES

Após a realização das etapas da Análise Fatorial, o pesquisador poderá aproveitar os resultados desta pesquisa para estudos futuros semelhantes, principalmente se o objetivo foi a redução de dados, como a obtenção de variáveis latentes através das variáveis originais. Estas variáveis podem ser submetidas a estudos posteriores a fim de simplificar, diminuir custo ou reduzir tempo em estudos similares.

O trabalho também contribuirá para direcionar futuras pesquisas que buscam divulgar a análise de dados nas empresas. Como por exemplo, orientar trabalhos que pretendem investigar a respeito do mercado de trabalho para estatísticos e para quem necessitar utilizar a análise fatorial com o critério de rotação Varimax, em seus trabalhos acadêmico.

Sugestões para trabalhos futuros que se assemelham a este:

- Identificar os problemas que as empresas buscam solucionar internamente ao contratarem serviços de análises de dados.
- Verificar se as empresas conhecem e estão acompanhando as tendências de mercado (business-intelligence/business-analytics) que utilizam a estatística nas tomadas de decisões.

REFERÊNCIAS BIBLIOGRÁFICAS

ABEP. **Associação Brasileira de Empresas de Pesquisa**, 2017. Disponível em: <<http://www.abep.org/historia-da-pesquisa>>. Acesso em: 01 Outubro 2017.

BELFIORE, P. P.; FÁVERO, L. P. L.; ANGELO, F. D. APLICAÇÃO DE TÉCNICAS ESTATÍSTICAS MULTIVARIADAS EM EMPRESAS DE OPERAÇÃO LOGÍSTICA NO BRASIL EM FUNÇÃO DE INDICADORES ECONÓMICO-FINANCEIROS. **REAd**, São Paulo, v. N°3, n. 51, p. 22, junho 2006.

CAREERCAST. **CareerCast**, 2017. Disponível em: <<http://www.careercast.com/jobs-rated/best-jobs-2017>>. Acesso em: 20 Setembro 2017.

DA SILVA, L. A.; MARQUES PERES, ; BOSCARIOLI, C. **Introdução a Mineração de dados com aplicações em R**. 3°. ed. Rio de Janeiro: Elsevier, 2016.

DUHIGG, C. **O poder do hábito**: Por que fazemos o que fazemos na vida e nos negócios. Rio de Janeiro: Objetiva, 2012.

EMPRESOMETRO. **Estatística do número de empresas abertas em Ji-Paraná/RO**, 2017. Disponível em: <<https://www.empresometro.com.br/Home/Estatisticas>>. Acesso em: 08 Setembro 2017.

GEM. **Monitor Global do Empreendedorismo**, 2015. Disponível em: <[http://www.bibliotecas.sebrae.com.br/chronus/ARQUIVOS_CHRONUS/bds/bds.nsf/c6de907fe0574c8ccb36328e24b2412e/\\$File/5904.pdf](http://www.bibliotecas.sebrae.com.br/chronus/ARQUIVOS_CHRONUS/bds/bds.nsf/c6de907fe0574c8ccb36328e24b2412e/$File/5904.pdf)>. Acesso em: 15 Maio 2017.

HAIR, J. F. J.; BLACK, W. C.; BABIN, B. J. **Análise Multivariada de dados**. São Paulo: Bookman, 2009.

IBGE. **Instituto Brasileiro Geografia e Estatística**, 2017. Disponível em: <<http://www.ibge.gov.br/home/estatistica/economia/microempresa/>>. Acesso em: 25 Junho 2017.

MANLY, B. F. J. **Métodos Estatísticos Multivariados Uma Introdução**. Porto Alegre: Bookman, 2008.

MAROCO, J.; MARQUES, T. G. Qual a fiabilidade do alfa de Cronbach? Questões antigas e soluções modernas? **Instituto Superior de Psicologia Aplicada** , Portugal, p. 86, jul. 2006.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2013.

PROOF. **PROOF**, 2015. Disponível em: <<http://www.proof.com.br/blog/business-intelligence-x-business-analytics/>>. Acesso em: 20 set. 2017.

PROVOST, F.; FAWCETT, T. **Data Science para Negócios**. Rio Janeiro: Alta Books, 2016.

RODRIGUES, ; BLATTMANN, U. Gestão da informação e a importância do uso de fontes de informação para geração de conhecimento. **Perspectivas em Ciência da Informação**, v. 19, n. 3, p. 4-29, Jul/Set 2014.

SEBRAE. **Portal SEBRAE**, 2017. Disponível em: <<http://www.sebrae.com.br/sites/PortalSebrae/>>. Acesso em: 28 Junho 2017.

SILVA, L. A. D.; PERES, S. M.; BOSCARIOLI, C. **Introdução a mineração de dados**: com aplicações em R. 3ª. ed. Rio de Janeiro: Elsevier, 2016.

APÊNDICES

APÊNDICE I – CARTA DE APRESENTAÇÃO

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO - TCLE BASEADO NAS DIRETRIZES CONTIDAS NA RESOLUÇÃO CNS N ° 466/2012,MS.

O senhor (a) esta sendo convidado (a) como voluntário (a) a participar da pesquisa: Validação de uma pesquisa realizada com empresários de Rondônia sobre a demanda de pesquisas de mercado nas empresas.

Esta pesquisa está sendo desenvolvida pela aluna Tuanny Rozeira Haverroth da Universidade Federal de Rondônia - UNIR, como parte de seu Trabalho de conclusão de curso TCC, sob a orientação do Professor DR. Nerio Aparecido Cardoso, com a finalidade de promover a pesquisa científica e disseminar este conhecimento para os profissionais da estatística, mostrando a importância do conhecimento científico na prática diária do estatístico nas empresas.

Solicitamos a sua colaboração para responder o questionário, como também sua autorização para apresentar os resultados deste estudo em eventos da área de estatística e administração e publicar em revista científica nacional e/ou internacional. Por ocasião da publicação dos resultados, sua identidade será mantida em sigilo absoluto. Os riscos envolvidos na pesquisa são mínimos, apenas poderá haver um desconforto pelo tempo usado para responder ao questionário.

Esclarecemos que sua participação no estudo é voluntária e, portanto, o (a) senhor (a) não é obrigado (a) a fornecer as informações e/ou colaborar com as atividades solicitadas pelos Pesquisadores caso decida não participar do estudo, ou resolver a qualquer momento desistir do mesmo, não sofrerá nenhum dano. Os pesquisadores estarão a sua disposição para qualquer esclarecimento que considere necessário em qualquer etapa da pesquisa.

Assinatura do pesquisador responsável

Considerando, que fui informado(a) dos objetivos e da relevância do estudo proposto, de como será minha participação, dos procedimentos e riscos decorrentes deste estudo, declaro o meu consentimento em participar da pesquisa como também concordo com os dados obtidos na investigação sejam utilizados para fins científicos (divulgação em eventos e publicações). Estou ciente que receberei uma via desse documento.

Ji-Paraná, ____ de _____ de 2017.

Assinatura do participante

Contato do Pesquisador Responsável, caso necessite de maiores informações sobre o presente estudo, favor entrar em contato com o pesquisador Dr. Nerio Aparecido Cardoso, e-mail: neriocardoso@hotmail.com.

APÊNDICE II – INSTRUMENTO DE REGISTO



FUNDAÇÃO UNIVERSIDADE FEDERAL DE RONDÔNIA - UNIR
CAMPUS JI-PARANÁ - Departamento de Matemática e Estatística



DADOS DE IDENTIFICAÇÃO			
Nome:		Empresa:	
Cargo:		Telefone:	
E-mail:			
COLETA DE DADOS			
1. Seguimento da empresa: a. () Comércio b. () Indústria c. () Serviço d. () _____			
2. Faturamento anual da empresa em média: R\$ _____			
3. Estrutura de Pessoal (Nº de trabalhadores) :			
###Breve explicação sobre o que é Análises de dados internos ou externos.			
4. Você já contratou serviços de Análises de dados internos ou externos? Justifique sua resposta. a. () Sim. b. () Não R _____			
5. Teria interesse em contratar? Justifique sua resposta. a. () Sim. b. () Não R _____			
6. Rotina da empresa: Responda de 1 a 5 sendo: [1] Nunca [2] Pouco [3] As vezes [4] Quase sempre [5] Sempre			
Nº	Afirmações	Avaliação	Desde quando
A	A empresa tem registro cadastral dos clientes		
B	A empresa registra a quantidade e itens de produtos que compra		
C	A empresa registra a quantidade e itens de produtos que vende		
D	A empresa utiliza os dados dos registros internos para extrair informações para melhorar o planejamento		
E	A empresa disponibiliza algum canal de sugestão/reclamação para os clientes		
7. Avalie os benefícios de se investir em Análises de Dados internos e externos. Para você qual o grau de importância nas seguintes afirmações. Responda de 1 a 5 sendo: [1] Nenhuma benefício [5] Muito benefício.			
Nº	Afirmações	Avaliação	
A	Análises de dados ajuda a Conhecer perfil de clientes		
B	Entender a necessidade do cliente		
C	Auxiliar na tomada de decisões		
D	Ajuda na hora de comprar mais mercadoria		
E	Criar promoções		

F	Conhecer a concorrência	
G	Conhecer falhas internas da empresa	
H	Aumenta os lucros	
I	Melhorar gestão	
J	Melhor Direcionar investimentos para propaganda	
8. Avalie as dificuldades de se investir em Análises de Dados internos e externos. Para você qual o grau de importância nas seguintes afirmações? Responda de 1 a 5 sendo: [1] Nenhuma dificuldade [5] Muita dificuldade.		
Nº	Afirmações	Avaliação
A	Não fornecem informações importantes	
B	Nos resultados não terá nada que eu já não saiba sobre minha empresa e meus clientes	
C	Não consigo colocar em prática as informações extraídas das análises.	
D	Os valores são altos para contratar esse tipo de serviço	
E	Não conheço quem faça esse trabalho	
F	Eu mesmo faço minhas análises de forma empírica	
G	Temos funcionários capacitados para fazer as análises	
H	Difícil contratar um profissional ou empresa dessa área	
I	Não gosto de ser avaliado	
J	Não gosto de tecnologia	

APÊNDICE III – ROTINA DE ANÁLISE EM AMBIENTE R

ANÁLISE FATORIAL COM ESTIMATIVA COMPONENTE PRINCIPAL E

ROTAÇÃO VARIMAX

Pacotes de funções do R utilizados na análise

```
library(mvShapiroTest)
library(psy)
library(car)
library(mvtnorm)
library(corrgram)
library(psych)
library(GPArotation)
library(corrplot)
library(fmsb)
library(ggplot2)
library(forcats)
library(ggcorrplot)
library(tseries)
library(gridExtra)
```

Carregando os dados

```
TCCC
```

#Análise descritiva

```
sumário(TCCC)
```

#Teste Cronbach - Resultados declarados na página 39.

```
cronbach(TCCC)
dim(TCCC)
196/25
```

#Teste Normalidade Shapiro-Wilk - Resultados declarados na página 39, Tabela 9.

#- Teste Normalidade Shapiro-Wilk Univariado – Resultados declarados na página

```
Shapiro_Wilk <- uniNorm(TCCC, type = "SW", desc = TRUE)
```

```
Shapiro_Wilk
```

Teste Normalidade Shapiro-Wilk Multivariado

```
result1 <- mshapiro.test(t(TCCC))
```

```
result1
```

Matriz de Correlação

```
cor(TCCC)
```

Gráfico de correlação das 25 variáveis - Resultados declarados na página 40, Figura 3

```
corrplot(cor(TCCC), order = "hclust", tl.col='black', tl.cex=.75)
```

Teste de correlação de Bartlett- Resultados declarados na página 40.

```
cortest.bartlett(TCCC, n = 196)
```

#Teste de KMO e MSA – Resultados declarados na página 41, Tabela 10.
KMO(TCCC)

#Retirar do banco de dados Var 19, Var 21, Var 23.
TCC_2 <- TCCC[, c(-19, -21,-23)]

#####Refazendo os testes

Gráfico de correlação das 22 variáveis - Resultados declarados na página 42, Figura 4.
corrplot(cor(TCC_2), order = "hclust", tl.col='black', tl.cex=.75)

Teste de correlação de Bartlett - Resultados declarados na página 42.
cortest.bartlett(TCC_2, n = 196)

#Teste Cronbach - Resultados declarados na página 43.
cronbach(TCC_2)
dim(TCC_2)
196/22

#Teste de KMO e MAS - Resultados declarados na página 43, Tabela 11.
KMO(TCC_2)

#Gráfico para visualizar quantos fatores devemos usar - Resultados declarados na página #44, Figura 5.
scree.plot(TCC_2)

#Matriz não rotacionada I - Resultados declarados na página 45, Tabela 12
AF_ao_rotacionada <- principal(TCC_2, nfactors = 4, rotate = 'none')
AF_ao_rotacionada

Na matriz não rotacionada precisamos retirar mais variáveis e refazer todo o procedimento
TCC_3 <- TCC_2[, c(-4,-5,-6,-7,-11,-13,-15,-16,-19,-20)]
TCC_3

#####Refazer os Testes

Gráfico de correlação das 12 variáveis - Resultados declarados na página 46, Figura 6.
X11()
corrplot(cor(TCC_3), order = "hclust", tl.col='black', tl.cex=.75)

Teste de correlação de Bartlett - Resultados declarados na página 47.
cortest.bartlett(TCC_3, n = 196)

#Teste Cronbach - Resultados declarados na página 47.
cronbach(TCC_3)
dim(TCC_3)
196/12

#Teste de KMO e MSA Tabela 13 - Resultados declarados na página 47.

```
KMO(TCC_3)
```

#Gráfico para visualizar quantos fatores devemos usar para as 12 variáveis - Resultados declarados na página 47, Figura 7.

```
x11()
```

```
scree.plot(TCC_2)
```

#Matriz não rotacionada II – Resultados declarados na página 49, Tabela 14.

```
AF_ao_rotacionada <- principal(TCC_3, nfactors = 4, rotate = 'none')
```

```
AF_ao_rotacionada
```

Matriz Rotacionada - (Varimax) - Resultados declarados na página 50, Tabela 15.

```
AF_rotacionada <- principal(TCC_3, nfactors = 4, rotate = 'varimax')
```

```
AF_rotacionada
```

Diagrama da af rotacionada (estimativas dos fatores foi dado pelo método de ACP)

###GRÁFICO DOS FATORES - Resultados declarados na página 52, Figura 9.

```
x11()
```

```
fa.diagram(AF_rotacionada, digits=6,e.size=0.09,rsiz=0.3,side=9, main="Análise Fatorial  
Através da Componente Principal", marg=c(0.5,0.5,2,0.5))
```

#Figura 10 - Resultados declarados nas páginas 55.

```
a<- ggplot(data = scores.Fa, aes(x = RC1, y = RC2, color = TCCC$Setor)) +  
  geom_point(aes(shape = TCCC$Setor), size=5, alpha=0.8) +  
  labs(title = "a) Plano Fatorial 1-2 com a variável categórica (Setor de atuação)",  
        x = "Fator 1 (Benefícios): 24%",  
        y = "Fator 2 (Inovação): 17%")+  
  scale_colour_discrete("Setor")+  
  scale_shape_discrete("Setor")
```

```
c<- ggplot(data = scores.Fa, aes(x = RC1, y = RC3, colour = (TCCC$Setor))) +  
  geom_point(aes(shape= (TCCC$Setor)), size=5, alpha=0.8) +  
  labs(title = "b) Plano Fatorial 1-3 com a variável categórica (Setor de atuação)",  
        x = "Fator 1 (Benefícios): 24%",  
        y = "Fator 3 (Dificuldades): 16%")+  
  scale_colour_discrete("Setor")+  
  scale_shape_discrete("Setor")
```

```
e<-ggplot(data = scores.Fa, aes(x = RC1, y = RC4, colour = (TCCC$Setor))) +  
  geom_point(aes(shape= (TCCC$Setor)), size=5, alpha=0.8) +  
  labs(title = "c) Plano Fatorial 1-4 com a variável categórica (Setor de atuação)",  
        x = "Fator 1 (Benefícios): 24%",  
        y = "Fator 4 (Organização): 14%")+  
  scale_colour_discrete("Setor")+  
  scale_shape_discrete("Setor")
```

```
x11()
```

```
grid.arrange(a,c,e)
```

#Figura 11 - Resultados declarados nas páginas 56.

```
b<- ggplot(data = scores.Fa, aes(x = RC1, y = RC2, colour = TCCC$JC)) +  
  geom_point(aes(shape = TCCC$JC), size=5, alpha=1) +  
  labs(title = "a) Plano Fatorial 1-2 com a variável categórica (Já contratou)",  
        x = "Fator 1 (Benefícios): 24% ",  
        y = "Fator 2 (Inovação): 17%")+  
  scale_colour_discrete("Já contratou")+  
  scale_shape_discrete("Já contratou")  
  
d<-ggplot(data = scores.Fa, aes(x = RC1, y = RC3, colour = (TCCC$JC))) +  
  geom_point(aes(shape= (TCCC$JC)),size=5, alpha=0.8) +  
  labs(title = "b) Plano Fatorial 1-3 com a variável categórica (Já contratou)",  
        x = "Fator 1 (Benefícios): 24% ",  
        y = "Fator 3 (Dificuldades): 16%")+  
  scale_colour_discrete("Já contratou")+  
  scale_shape_discrete("Já contratou")  
  
f<-ggplot(data = scores.Fa, aes(x = RC1, y = RC4, colour = (TCCC$JC))) +  
  geom_point(aes(shape= (TCCC$JC)), size=5, alpha=0.8) +  
  labs(title = "c) Plano Fatorial 1-4 com a variável categórica (Já contratou)",  
        x = "Fator 1 (Benefícios): 24% ",  
        y = "Fator 4 (Organização): 14%")+  
  scale_colour_discrete("Já contratou")+  
  scale_shape_discrete("Já contratou")  
  
x11()  
grid.arrange(b,d,f)
```